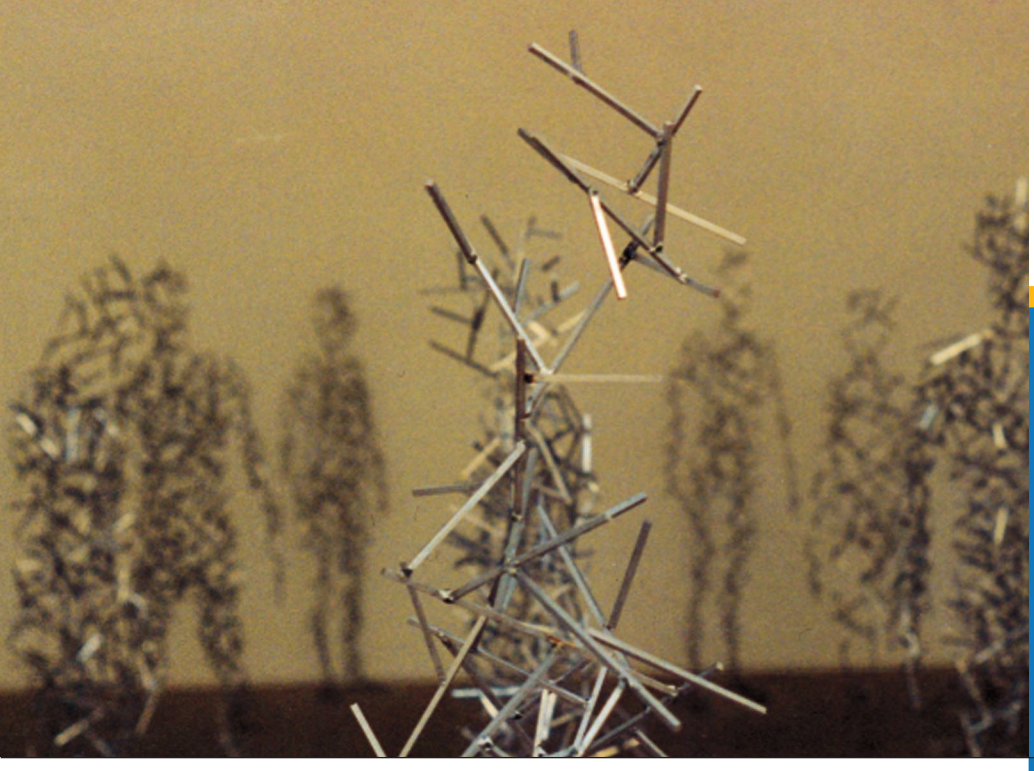


William Hennah

Genetics of Schizophrenia: The 1q42 Locus in Finnish Families

Publications of the National Public Health Institute A 22/2005

Department of Molecular Medicine,
National Public Health Institute Helsinki, Finland
and
Department of Medical Genetics,
University of Helsinki, Finland

William Hennah

GENETICS OF SCHIZOPHRENIA: THE 1Q42 LOCUS IN FINNISH FAMILIES

ACADEMIC DISSERTATION

To be presented with the permission of the Faculty of Medicine, University of Helsinki, for public examination in Lecture Hall 3, Biomedicum, on December 1st, at 12 noon.

National Public Health Institute, Helsinki, Finland

and

Department of Medical Genetics, University of Helsinki, Finland

Helsinki 2005

Publications of the National Public Health Institute
KTL A22 / 2005

Julkaisija-Utgivare-Publisher

Kansanterveyslaitos (KTL)
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 474 41, telefax (09) 4744 8408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 474 41, telefax (09) 4744 8408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 474 41, telefax +358 9 4744 8408

ISBN 951-740-571-5
ISSN 0359-3584
ISBN 951-740-572-3 (pdf)
ISSN 1458-6290 (pdf)

Helsinki University Biomedical Dissertations No. 70
ISSN 1457-8433
http://ethesis.helsinki.fi

Cover picture: William Hennah

Edita
Helsinki 2005

Supervised by

Academy Professor Leena Peltonen-Palotie
National Public Health Institute
Helsinki, Finland

Docent Tiina Paunio
National Public Health Institute
Helsinki, Finland

Reviewed by

Professor Hasse Karlsson
Univeristy of Turku
Turku, Finland

Docent Maija Wessman
University of Helsinki
Helsinki, Finland

Opponent

Professor David Porteous
Department of Medical Sciences
University of Edinburgh

“Voimallinen tahto vie miehen läpi harmaan kiven”
(A man’s will can take him through stone) – Aleksis Kivi

William Hennah, Genetics of Schizophrenia: The 1q42 Locus in Finnish Families
Publications of the National Public Health Insitute, A22/2005, 92 Pages
ISBN 951-740-571-5; 951-740-572-3 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/suomi/julkaisut/julkaisusarjat/kansanterveyslaitoksen_julkaisuja_a/

Abstract

Schizophrenia is a major mental illness which for a long time has been considered to have both genetic and environmental aspects to its aetiology. In recent years, genetic studies of schizophrenia have suggested that the long arm of chromosome 1 is a major candidate region for susceptibility loci for this complex disorder. With a number of research groups having reported some evidence of linkage or association at 1q21-24, 1q32.2-41, or 1q42. Although the data from all the implied chromosomal loci for schizophrenia are not consistent, the findings for chromosome 1q are amongst the most replicated across various study populations.

Initially 1q42 was identified as a potential site for a schizophrenia susceptibility locus in a large Scottish family, containing a wide spectrum of major mental illnesses, where a balanced (1;11)(q42.1;q14.3) translocation was found to co-segregate with schizophrenia and other related psychiatric disorders. Two novel overlapping genes located on the opposing DNA strands were found to be directly disrupted by this translocation, and were subsequently named Disrupted in Schizophrenia 1 and 2 (*DISC1* and *DISC2*). Linkage analysis carried out in Finnish schizophrenia families confirmed the 1q42 locus and the *DISC* genes as good positional candidates for susceptibility to schizophrenia, when a lod_{max} of 3.2 was observed for a microsatellite located intragenic of *DISC1*.

In the studies performed here the initial linkage finding on 1q42 was verified through its replication in an independent sample of families from Finland. This locus was further investigated using SNPs and their haplotypes, monitoring for transmission disequilibrium to affected individuals in the entire population of 458 Finnish families. This analysis highlighted four restricted genomic regions that significantly associate with schizophrenia. Amoung these, the most robust finding is the associated allelic haplotype that spans from intron 1 to exon 2 of the *DISC1* gene (HEP3 haplotype), and was found to be significantly under-transmitted to affected females.

The haplotype was used to further dissect the role of *DISC1* in the aetiology of schizophrenia, and it was found that the association was mainly to qualitative trait-components representing delusions, hallucinations, and negative symptoms of schizophrenia. In using quantitative traits that represent endophenotypes of

schizophrenia, it was found that the haplotype associated mainly to visual working memory functions. This association was between HEP3 and poorer performance in visual working memory, particularly in males, leading to the analysis of a control sample, representative of the general Finnish population. This enabled the re-evaluation of the previously observed under-transmission to affected females as an epiphenomenon caused by the higher frequency of HEP3 in the Finnish families ascertained for schizophrenia. Consequently, the data points to HEP3 representing a variation that is associated with higher risk to schizophrenia in males and suggests that it functions through a mechanism affecting visual working memory.

Since schizophrenia is known to be a polygenic disorder it was of interest to analyse which genomic regions may be linked to schizophrenia when simultaneously accounting for the effect of *DISC1*. A genome scan was performed on two HEP3 stratified samples, with three linkage peaks (lod > 3) being observed. One was the original 1q42 locus containing *DISC1*, while the other two represented a region previously significantly linked to schizophrenia (10q21) and a locus (16p13) containing a *DISC1* interacting gene (nuclear distribution gene E homolog 1, *NDE1*). Among the peaks displaying suggestive lod scores, six loci were identified in regions previously linked to schizophrenia, including the loci for the candidate genes dysbindin (*DTNBP1)*, neuregulin 1 (*NRG1)*, glutamate receptor metabotrophic 3 (*GRM3)*, and reelin (*RELN)*. Of main interest though was the identification of a locus containing a *DISC1* interacting gene, *NDE1*, as it highlights the pathway these two genes act along for a role in schizophrenia. *NDE1* was further analysed and found to associate to schizophrenia, the risk allele being significantly associated with affected females. Suggesting that a *DISC1* "pathway" is involved in the aetiology of schizophrenia in the Finnish population

Keywords: Schizophrenia, Linkage, Association, *DISC1*, *NDE1*

CONTENTS

Abbreviations

AF	all Finland except the isolate sample set
AKT1	v-*akt* murine thymoma viral oncogene homolog 1
ASP	affected sibling pair
cAMP	cyclic adenosine monophosphate
CAPON	C-terminal PDZ domain ligand of neuronal nitric
CHRNA7	cholinergic receptor, nicotinic, alpha polypeptide 7
cM	centiMorgan
COMT	catechol-O-methyltransferase
CVLT	California verbal learning test
DAO	D amino acid oxidase
df	degrees of freedom
DISC	disrupted in schizophrenia
DNA	deoxyribonucleic acid
DSM	diagnostic and statistical manual of mental disorders
DTNBP1	dystobrevin binding protein 1 (dysbindin)
EDTA	ethylenediaminetetraacetic acid
EPSIN4	epsin 4
EST	expressed sequence tag
FEZ1	fasciculation and elongation protein zeta 1
G30	hypothetical protein *G30*
G72/DAOA	D-amino acid oxidase activator
GABA	gamma-aminobutyric acid
GRM3	glutamate receptor, metabotropic 3
HEP	author's arbitrary naming of identified haplotypes
IBD	identical by descent
ICD	international classification of diseases

IS	isolate sample set
kb	kilobase
LC	liability class
LD	linkage disequilibrium
LIS1	platelet-activating factor acetylhydrolase
lod	logarithm of odds
lod_{max}	maximum logarithm of odds
MCMC	Markov chain Monte Carlo
MRI	magnetic resonance imaging
n	number of samples
NA	not applicable
NDE1	nuclear distribution gene E homolog 1
NDEL1	nuclear distribution gene E homolog like 1
NRG1	neuregulin 1
OCCPI	operational criteria checklist for psychotic illness
p	p-value
PDE4B/D	phosphodiesterase 4B/4D, cAMP-specific
PPP3CC	protein phosphatase 3, catalytic
PRODH	proline dehydrogenase 1
RELN	reelin
RGS4	regulator of G-protein signalling 4
RNA	ribonucleic acid
SNP	single nucleotide polymorphism
TRAX	translin-associated factor X
WAIS-R	Wechsler adult intelligence scale - revised
WMS-R	Wechsler memory scale - revised

List of original publications

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals.

I Ekelund J, Hennah W, Hiekkalinna T, Parker A, Meyer J, Lönnqvist J, Peltonen L. (2004) Replication of 1q42 Linkage in Finnish Schizophrenia Pedigrees. **Mol. Psychiatry. 9(11):1037-1041**

II Hennah W, Varilo T, Kestilä M, Paunio T, Arajärvi R, Haukka J, Parker A, Martin R, Levitzky S, Partonen T, Meyer J, Lönnqvist J, Peltonen L, Ekelund J. (2003) Haplotype Transmission Analysis Provides Evidence of Association for *DISC1* to Schizophrenia and Suggests Sex-Dependent Effects. **Hum. Mol. Genet. 12(23):3151-3159**

III Hennah W, Tuulio-Henriksson A, Paunio T, Ekelund J, Varilo T, Partonen T, Cannon TD, Lönnqvist J, Peltonen L. (2005) A Haplotype within the *DISC1* Gene is Associated with Visual Memory Functions in Families with a High Density of Schizophrenia. **Mol. Psychiatry. in press**

IV Hennah W, Tomppo L, Hiekkalinna T, Ekelund J, Tuulio-Henriksson A, Palo OM, Kilpinen H, Kestilä M, Silander K, Varilo T, Paunio T, Terwilliger JD, Lönnqvist J, Peltonen L. (2005) Families with the Risk Allele of *DISC1* Reveal a Link Between Schizophrenia and Another Component of the Same Pathway, *NDE1*. **Submitted**

1 Introduction

Schizophrenia is a severe mental disorder affecting approximately 1% of the worldwide population. The disorder is considered to be partly genetic in aetiology, with family[11], twin[12], and adoption[15] studies suggesting a considerable high heritability, estimated to be as high as 83% based on the analysis of a Finnish twin cohort[12].

Finland is one of the best characterized populations considered as relatively isolated and the established frequency of known disease mutations reflects the history of multiple bottlenecks[16]. This population history and the presence of extensive genealogical records make the Finnish population valuable for studies of genetic disorders. To date 36 diseases have been identified as enriched in the population, and are often referred to as the Finnish disease heritage[17,18]. It is hypothesised that the same events in the population history which led to this enrichment could also result in an increased homogeneity of the disease alleles underlying more complex disorders like schizophrenia[18-20].

The research presented here used three advances that have led to huge leaps in the genetic analysis of complex traits. The completion of the Human Genome Project has produced genome-wide information that can be analysed thoroughly *in silico*, providing tools for efficient genome-wide analysis and directing the further analyses to be performed. Improvements in high throughput genotyping now means that typing large numbers of markers, in samples of large enough size to give the power to detect association, is no longer cost prohibitive. Lastly, there have been advances in the statistical programs used for analysing such large amounts of data robustly. These three were applied to the analysis of the 1q genomic locus in a sample of Finnish families with a high density of schizophrenia, which was established to contain potential susceptibility genes through prior genetic analysis of schizophrenia. Furthermore, these techniques were applied to the continuation of the analysis, by dissecting the role of the 16p13 region in schizophrenia, a locus highlighted by linkage analysis conditioned on 1q42.

2 Review of the literature

2.1 Overview of schizophrenia

2.1.1 Diagnosis

Schizophrenia is a heterogeneous disorder identified on the basis of a pattern of abnormal behaviours, characterised by false beliefs and abnormal perceptions, which are reflective of the differentiation between inner preconceptions and expectations from external stimuli. However, a number of other symptoms can also present with diagnosis being dependent on the quantity, severity and duration of these symptoms.

Hippocrates (460-377B.C.) described epilepsy, mania, and depression, which he correctly attributed to the brain. However, few ancient texts are found to describe schizophrenia, yet upon adequate description of the disorder in 1809 the disease became more visible over the western world and numbers of sufferers increased rapidly for a hundred years[21]. Nowadays schizophrenia is one of the best known and most common forms of mental illness, yet its true origins remain a mystery.

These first adequate descriptions of schizophrenia were made independently in England and France, by John Haslam (1764-1844) and Philippe Pinel (1745-1826) respectively. Yet, it is Emil Kraepelin (1856-1926) who is so often credited, due to his definitive work in categorising the disorder and in naming it *dementia praecox*. Most of Kraeplein's clinical analyses of schizophrenia still stand as the descriptive terms used today[21].

Modern day schizophrenia can currently be diagnosed according to two methods, the Diagnostic and Statistical Manual for Mental Disorders currently in its forth edition (DSM-IV)[28] (Table 1), and the International Classification of Diseases, now in its tenth edition (ICD-10)[29]. The two classification systems do concur and essentially only differ in their semantics[21]. In Europe both methods can be used for clinical and research work, yet the ICD-10 predominates in the clinical field and the DSM-IV in the research field.

Table 1

DSM-IV diagnostic criteria for schizophrenia
A. Characteristic symptoms: Two or more of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated): 1. Delusions 2. Hallucinations 3. Disorganized speech 4. Grossly disorganized or catatonic behaviour 5. Negative symptoms Only one Criterion A symptom is required if delusions are bizarre or hallucinations consist of a voice keeping up a running commentary on the person's behaviour or thoughts, or two or more voices are conversing with each other. **B.** Social/occupational dysfunction: for a significant portion of the time since the onset of the disturbance, one or more major areas of functioning such as work, interpersonal relations, or self-care are markedly below the level achieved prior to the onset (or, when the onset is in childhood or adolescence, failure to achieve the expected level). **C.** Duration: Continuous signs of the disturbance persist for at least 6 months, of which at least one month should be of symptoms that meet Criterion A. The 6 months may include periods of prodromal and residual symptoms. **D.** Schizoaffective and mood disorder exclusion: Schizoaffective disorder and mood disorder with psychotic features have been ruled out because either no major depressive, manic, or mixed episodes have occurred concurrently with the active-phase symptoms, or if mood episodes have occurred during active-phase symptoms, their total duration has been brief relative to the active and residual periods. **E.** Substance/general medical condition exclusion: The disturbance is not due to the direct physiological effects of a substance or a general medical condition. **F.** Relationship to a pervasive developmental disorder: if there is a history of autistic disorder or another pervasive developmental disorder, the additional diagnosis of schizophrenia is made only if prominent delusions or hallucinations are also present for at least a month (or less if successfully treated).

There are a number of other psychiatric disorders that present with partially similar symptoms to schizophrenia[30], and it is common to observe these disorders in families where a schizophrenia affected individual has already been identified. This leads to the hypothesis that these disorder symptoms are in a

continuum, rather than being distinct disorders, suggesting that there will be some common genes underlying psychoses that fall into separate diagnostic classes of the classification system.

2.1.2 Epidemiology

To understand a disorder like schizophrenia it is necessary to quantify it in a number of ways that allow for the identification of the natural characteristics of the disorder. Primarily, these are the morbid risk (likelihood of someone suffering an episode of schizophrenia during their life), incidence (number of new cases of schizophrenia usually given in a period of one year and in a population of 100,000), and prevalence (total number of living people who have been diagnosed with the disorder). Although schizophrenia is a global phenomenon such measures differ between the population samples studied, for example the prevalence of the disorder can range from 0.3 per 1,000[35] to 22 per 1,000[36]. Such variations can be observed between countries, municipalities or districts, and between urban and rural areas[37]. However, it is widely stated that schizophrenia has a global morbid risk of about 1%, and while individuals with schizophrenia generally have a decreased reproduction rate[38] the prevalence seems to be relatively stable over time. One such hypothesis for this stability is that the genetic variations involved in the aetiology are relatively common as they may provide an advantage to the individual (for review see[39]).

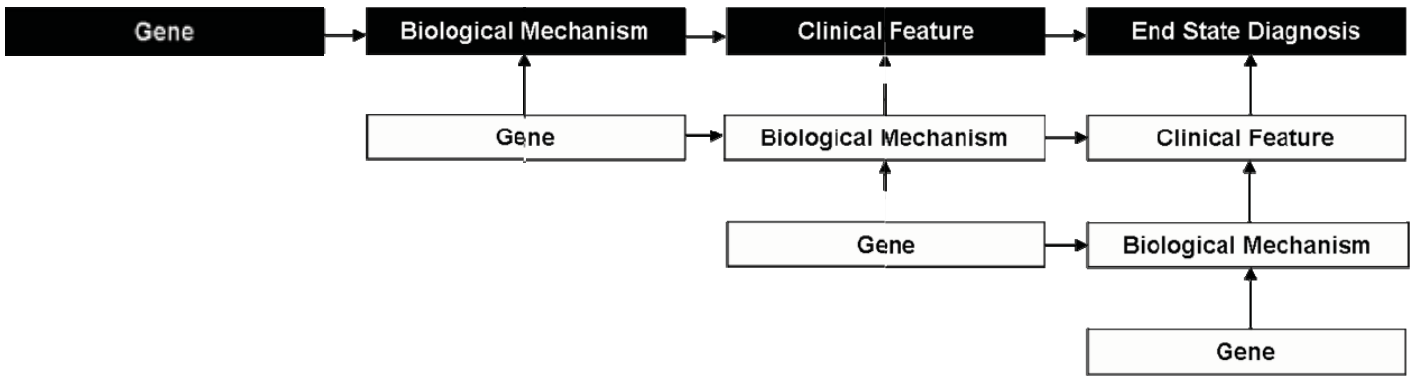
On the whole schizophrenia is equally prevalent in men and women, although some studies do report a higher prevalence in males[36,45,46]. There are some sex differences with regard to the prognosis of the disorder with men on average having an earlier onset and more severe course to the disorder. The peak for men is in the early twenties, while for women it is in the mid to late twenties with a second peak around the time of menopause[47].

Family[11], twin[12] and adoption[15] studies have shown that there is a highly genetic aetiology to schizophrenia. This genetic epidemiology is discussed in detail later. Despite the high genetic risk there is also a strong risk from environmental factors, with risks caused by genetic and environmental interactions often being included with the genetic risk. Numerous environmental risk factors have been identified for schizophrenia, with the most widely regarded including obstetric complications leading to foetal hypoxia[52,53], infections during pregnancy[55-57], urban birth and upbringing[58], and heavy cannabis use during adolescence[59].

2.1.3 Clinical features

Schizophrenia is a complex phenotype, consisting of many symptoms, some of which do not need to be present in order to meet the diagnostic criteria. Therefore, the clinical presentation of the disorder within a population of affected individuals is highly heterogeneous. This can obscure the potential effect of some loci when analysis is performed to search for linkage or association with schizophrenia in such populations, leading to many false negative results being observed. This can also be due to the fact that the effect of genetic variation can be obscured by other biological mechanisms that are acting on the disorder (Figure 1).

Figure 1
An oversimplified schematic presentation of the processes that lead from genes to the end-state diagnosis in a complex genetic disorder. In actuality it is likely that many genes, some of which will interact with each other, will act on each biological mechanism, and that these mechanisms potentially interact together and provide liability to possibly many of the clinical features that make up the end state diagnosis. Additionally, the environment could provide further risks, or interact with any of the stages presented here.



One approach to address the complexity of the clinical phenotype, not necessarily reflecting the biological background of the disease, is to use clinical features as intermediate phenotypes, whose relationship to specific molecular pathways might be more direct. These can be either trait components or endophenotypes, which only differ in their relation to the disorder of interest. Trait components are generally dependent on the disorder and serve as an aid to making the studied phenotype more homogenous. For example a particular symptom of schizophrenia, like auditory hallucinations, can be used. This trait would not generally be present if the individual was not affected, yet not all individuals with schizophrenia present that trait. Importantly, even if a person without the disorder of interest has a symptom that is used as a trait component they will not be classified as affected if the main diagnosis is not present. In contrast an endophenotype is an independent trait that is associated with the disorder. In the case of schizophrenia working memory can be used as an example of an endophenotype. Everyone has a degree of working memory, yet it has been shown that individuals with schizophrenia on average perform worse on tasks requiring this function. The purpose to using such traits is that their molecular background is likely to be more straightforward, with less genetic heterogeneity contributing to the variation in the trait, when compared to the diagnostic phenotype of schizophrenia[65,66]. This is summarised in the five criteria for what constitutes an endophenotype; 1) it is associated with illness in the population, 2) it is heritable, 3) primarily state-independent, 4) co-segregates with illness in families, and 5) unaffected family members of affected individuals have a higher rate of the endophenotype than the general population[65]. In order to be provide more power to genetic analysis a sixth criterion can be added that specifies that the endophenotype be a quantitative measure.

For schizophrenia trait components can be derived, for example, from the Operational Criteria Checklist for Psychotic Illness (OCCPI), as used in II, which is a series of 90 items representing life history and symptomatology[69,70] which can be systematically checked if a feature is present, creating an individual profile for each presenting proband. Prominent candidate endophenotypes are measures of the structure and function of the brain. These can be traits representative of cognitive functions as used in III and IV, but can also be derived from electrophysiological and structural and functional imaging of the brain. In a practical sense, it can be hypothesized that the use of more informative quantitative traits, if truly associated with clinical vulnerability and the studied gene variations, has the potential to display stronger association with the genetic markers, than the end-state diagnosis.

Endophenotypes of schizophrenia have previously proven to be successful in determining more about the relation between the finding of the cholinergic receptor, nicotinic, alpha polypeptide 7 (*CHRNA7*) gene association and schizophrenia. When it was observed that the gene associates to deficits in the mechanism that underlies the subconscious ability to gradually gate the response to a repeated stimulus[71], sensory gating. This is represented by the characteristics of the P50 evoked potential which normally decreases upon reoccurrence of a stimulus, yet stays at its original intensity in the brains of individuals with schizophrenia and their unaffected relatives[72].

2.2 Genetics of schizophrenia

2.2.1 General aspects

It has long been thought that schizophrenia has a partly genetic aetiology, but it was only through family, twin and adoption studies that this hypothesis gained credibility.

Large numbers of studies have shown that schizophrenia runs in families, as the morbid risk of developing schizophrenia increases with the greater genetic relatedness to an individual with the disorder (Table 2)[73]. It has also been shown that the number of affected siblings an individual has increases their risk of developing schizophrenia. When an epidemiological study was carried out in Finland, it was observed that the morbid risk of schizophrenia to individuals was 4.2%, 6.4%, and 8.7% given 1, 2, or 3, affected siblings respectively[36]. However, family evidence is not enough to prove a genetic aetiology to a disorder as families generally share much of their environment as well.

Twin studies go some way to addressing the matter of elucidating the contribution of genes to the disorder. With a higher concordance of schizophrenia between monozygotic twins (48%) than between dizygotic twins (17%)[12,74] suggesting a clear contribution to the aetiology from genetics. Although this method has some drawbacks, the method can not account for differences in the way mono- and di-zygotic twins are treated similarly, it can not separate the effects of genetics from those of a shared *in utero* environment, and it can not account for gene expression differences between the twins.

To better account for the environmental issues that are involved in the aetiology of schizophrenia it is necessary to perform adoption studies. Such studies have found that there was increased risk of schizophrenia in the biological relatives of an adopted affected individual yet no increase in risk was observed for the adoptive relatives[15,75,76].

Twin studies have also played a role in determining the heritability of schizophrenia, and thereby stating what fraction of the disorder can be said to be due to genetic factors. In a Finnish study, such an estimate of heritability was made to be as high as 83%, with the remaining 17% being due to non-shared environmental factors[12], such as cannabis smoking by one twin and not the other.

Table 2
Morbid risk of schizophrenia depending on genetic relatedness to an affected individual, adapted from Tsuang 2000[73].

Relationship	Shared Genes	Morbid Risk (%)
General Population	NA	1
Spouse of Patients	NA	2
Third-Degree Relatives	12.5	
First Cousins		2
Second-Degree Relatives	25	
Uncles/Aunts		2
Nieces/Nephews		4
Grandchildren		5
Half-Siblings		6
First-Degree Relatives	50	
Parents		6
Siblings		9
Children		13
Sibling with 1 affected parent		17
Dizygotic twin		17
Monozygotic twin	100	48
Children with 2 affected parents	100	46

NA, not applicable

2.2.2 Previous genetic findings for schizophrenia

Genetic studies of schizophrenia have been carried out for many decades, but it has only been since the turn of the millennium that consistent linkage findings for schizophrenia have been identified and independently replicated, giving a solid basis for further analysis of the regions with regard to identifying hypothetical candidate genes. Such linkage results came from genome-wide scans, and through the identification of chromosomal rearrangements in large pedigrees, and have led to several consensus loci being identified, many of which have had potential predisposing genes identified at or close to the linked regions: 1q21-22 (regulator of G-protein signalling 4, *RGS4*; C-terminal PDZ

domain of neuronal nitric, *CAPON*)[77-80], 1q42 (disrupted in schizophrenia 1, *DISC1*)[3-9,81,82], 5q (gamma-aminobutyric acid, *GABA,* receptor cluster and *EPSIN 4*)[14,80,83-88], 6p23 (dysbindin, *DTNBP1*)[22-26,89-93], 8p22-p11 (neuregulin 1, *NRG1*)[41-43,91,94], 13q32 (D-amino acid oxidase activator, *G72/DAOA;* hypothetical protein G30, *G30*)[41,54,95,96], and 22q11-q13 (catchecol-O-methyltransferase, *COMT*; praline dehydrogenase 1, *PRODH*)[97-104]. The numerous inconsistent linkage findings for schizophrenia are thought to be due to the heterogeneity of the disorder and the diversity of the populations being analysed. However, a recent meta-analysis of genome scans suggests that some loci may contribute to schizophrenia in many or most populations, with loci whose effects are strongly detected in a unique population potentially being relevant to other populations[2]. This is coming to light more and more as a number of these genes are replicated for association in independent samples from around the world. The most consistently replicated so far have been the *NRG1* and *DTNBP1* genes. In addition to these genes identified by consistent linkage peaks many genes are being identified as associated to the disorder by being identified as good candidates by other means, such as expression analysis (Table 3).

Table 3
Genetic evidence for each current candidate gene for schizophrenia and related disorders, updated and adapted from Harrison and Weinberger 2005[1].

		Strength of Evidence to Schizophrenia		
Gene	**Locus**	**Linkage**	**Association**	**Altered Expression**
RGS4	1q21-22	++++	+++	++
DISC1	1q42	++++	+++++	Not Known
EPSIN4	5q33	+++	+	Not Known
GABA receptors	5q34	+++	++	Not Known
DTNBP1	6p22	++++	+++++	++
GRM3	7q21-22	+	++	x
RELN	7q22	+	+	+
NRG1	8p12-21	++++	+++++	+
PPP3CC	8p21	++++	+	+
DAO	12q24	+	++	Not Known
G72/G30	13q32-34	++	++	Not Known
AKT1	14q22-32	+	+	++
CHRNA7	15q13-14	++	+	+++
COMT	22q11	+++	++	+
PRODH	22q11	+++	+	x

Strength of evidence is arbitrarily represented by a number of + signs, representative of the number of independent replications for that finding.

2.2.3 Chromosome 1q in schizophrenia

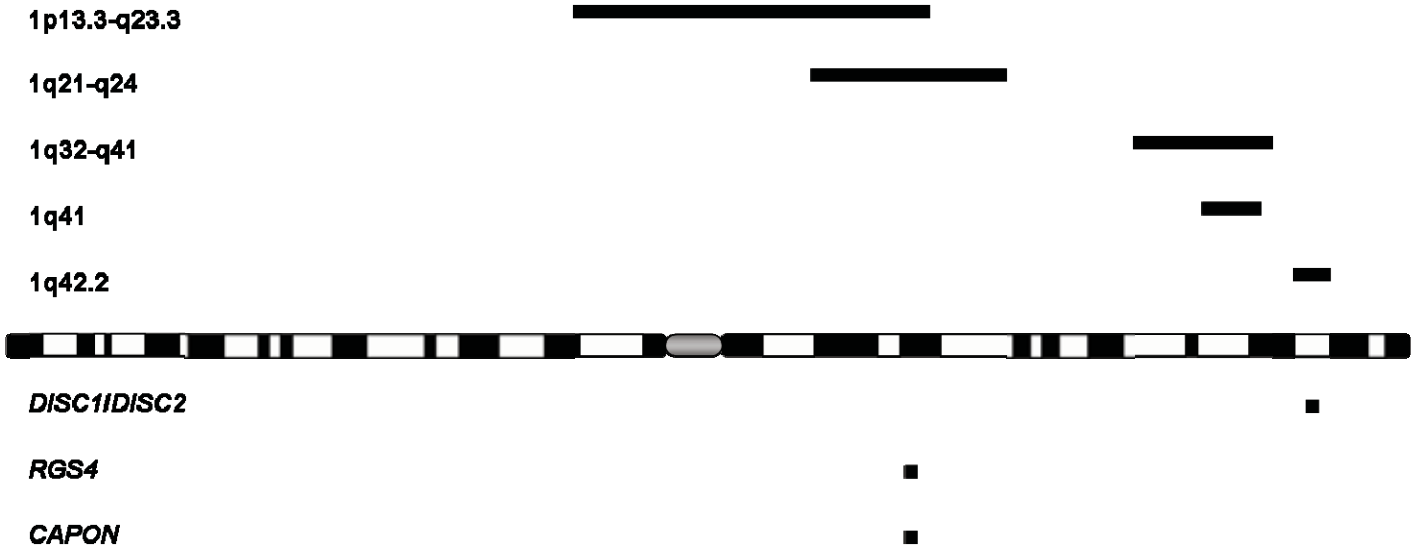
Chromosome 1q has emerged as one of the most likely regions in the genome for containing schizophrenia susceptibility loci (Figure 2). In Canadian families, a lod score of 6.5 has been reported on 1q21 between markers D1S1653 and D1S1679[77]. Some 10 cM telomeric from this location at 1q24, evidence of linkage between the trait and a polymorphic genetic marker displayed a logarithm of odds (lod; described in detail later) score of 3.2[80] in a British-Icelandic study sample around the marker D1S196. Further towards the 1q telomere, a haplotype of markers spanning the 1q32-q41 locus was identified by a study of an internal isolate population of Finland[105]. The markers from this haplotype were then genotyped in a case-control Finnish twin sample for schizophrenia with extensive neuropsychological data. The most telomeric of these markers at 1q41 displayed the strongest association to quantitative measures of visuospatial working memory and visual attention[106]. Also, in a family set ascertained for schizophrenia a genome-wide scan using a number of quantitative traits of cognitive functions observed some evidence of linkage (lod > 1.5) to verbal memory functions at the 1q41 locus[10]. However, it is a proximal location to this, 1q42, which has recently been developing as one of the most consistently replicated loci.

Initially 1q42 was identified as a potential site for a schizophrenia susceptibility locus in a large Scottish family, containing a wide spectrum of major mental illnesses, where a balanced (1;11)(q42.1;q14.3) translocation was found to co-segregate with schizophrenia and other related psychiatric disorders[3]. Two novel genes were found to be directly disrupted by this translocation, and were subsequently named Disrupted in Schizophrenia 1 and 2 (*DISC1* and *DISC2*)[107]. When other Scottish families ascertained for schizophrenia and bipolar disorder were used to monitor for association between common bi-allelic variations of one base pair found throughout the genome (single nucleotide polymorphisms or SNPs) in the *DISC* genes and the disorder, no significant association was observed[108]. Later, linkage findings within a Finnish schizophrenia nuclear family sample confirmed the 1q42 locus and the *DISC* genes as good positional candidates for susceptibility to schizophrenia, when a lod_{max} of 3.2 was observed for a microsatellite located within the *DISC1* gene[7].

Two large studies have failed to find support for any schizophrenia susceptibility locus on chromosome 1q. Initially, a multi-centre meta-analysis of all genome-wide scans of schizophrenia was performed, pooling linkage peaks rather than the original genotype information[2], chromosome 1q was not among the interesting loci, specifically the bin containing 1q42 represented the 46th bin among 120. However, a wide region spanning 1p13.3 to 1q23.3 was highlighted

in this study as a suggestive site for a schizophrenia locus (loci). The involvement of 1q in schizophrenia has also been challenged in a large multi-centre replication study with 984 schizophrenic sib-pairs. In total, 16 microsatellite markers, covering approximately 110 cM on 1q, were analyzed and no evidence of linkage was obtained[109]. Yet, the validity of the tests performed within these studies has been challenged[110,111]. One of the main weaknesses when using meta-analysis and multi-centre studies is that despite the increased sample size you also increase the genetic heterogeneity underlying the disorder, as such methods generally involve the use of many different sample populations of different ethnic origins. Such issues of genetic heterogeneity, incomplete penetrance and phenocopies are concerns that surround the whole of genetic research into complex disorders[112] and are thus exaggerated when using samples obtained from many different populations. Furthermore, meta-analyses also escalate many problems and biases of the initial studies, including how the diagnostic criteria were adapted, and the inheritance models used in the analysis.

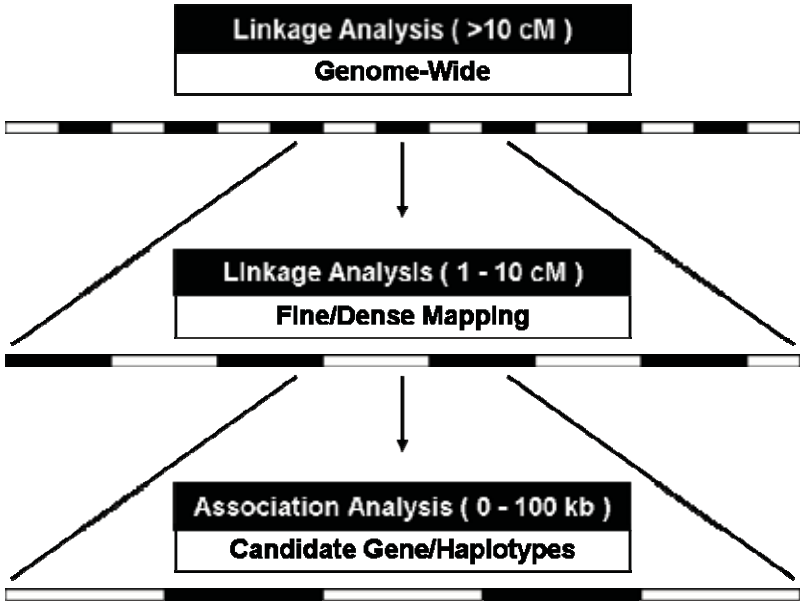
Figure 2
Diagram to show the relative locations of linkage (above) and association (below) observations on chromosome 1.



2.3 Strategies for mapping genes for complex disorders

In order to identify putative genes that underlie the aetiology of a complex disorder, a series of steps (Figure 3) are generally considered as the ideal way to proceed. Conventionally this starts with a whole genome scan that analyses hundreds of highly polymorphic genetic variations in the form of microsatellite markers from every region of the genome to identify large genomic regions that show co-segregation with the disorder in families or sibling pairs. Due to uncertainties in the parameters adapted in linkage analysis it is likely that a number of loci will be identified that are still very wide. Therefore, a dense or fine map is produced on the loci and genotyped in families in order to narrow down the critical DNA region. If linkage is constantly observed within the locus, analysing candidate genes is the next logical step, initially looking for association between intragenic and flanking SNPs and their haplotypes (described in full later). Certain priority is typically given in these analyses to genes biologically relevant for the studied disorder, such as neurotransmitter receptor genes in the case of schizophrenia. Genome-wide gene searches are often complimented with linkage and association studies of markers within or flanking functional candidate genes.

Figure 3
Illustration of the common three step strategy employed in genetic analysis to identify genomic regions of interest. White boxes provide the terms used to describe the analysis performed, and black boxes show the resolution provided at each step.



2.3.1 Statistical methods

2.3.1.1 Dichotomised trait analysis

2.3.1.1.1 Linkage analysis

Linkage analysis measures whether or not transmission of a certain region of a chromosome is non-random and dependent on the affection status of the disorder in question. In practice this analysis looks at the segregation of allelic variants of many microsatellite markers that are highly polymorphic exposing numerous alleles in the population. This is either done on a sparse genome-wide level, or on a fine/dense level of a region already of interest to the disorder. Then using a sample of families multiply affected by the disorder, statistical analyses are used to test if a particular allele of a particular marker segregates with the disorder at a greater frequency than by chance, in each individual family. This is repeated assuming a number of different recombination fractions between the marker and

the disease locus. Each marker is then given a logarithm of odds (lod) score for linkage for each family, which is then accumulated over the sample to produce an overall lod score for each recombination fraction[113,114]. The lod score is most often maximised over recombination fractions to produce a lod_{max} for each marker. For monogenic disorders the recombination fraction at which the lod_{max} occurs is indicative of the distance between the marker and the susceptibility gene, but in disorders with a complex genetic inheritance this property is largely ignored because the specified model for inheritance used in statistical analyses is often false or over simplified.

In the statistical analysis for significant deviation from chance, the high number of markers introduces the risk of type I errors (false positives). This high number has typically been around 300 when microsatellites have been used, but in the dawning age of genome-wide association analysis hundreds of thousands of SNP markers[115] are to be genotyped and analysed. It was necessary for guidelines to be made as to what constitutes a significant deviation from chance. With such guidelines for genome-wide analysis with microsatellites (Table 4) being set by Lander and Kruglyak in 1995, and are still generally adhered to today[116]. However, these will need revision to account for the increased scale of the new genome-wide association analyses.

Linkage analysis has been used for many years in the studies of complex genetic traits. Performing a genome-wide linkage scan has become a staple ingredient in all genetics departments worldwide. Yet such analysis comes with a heavy drawback, it is just the starting place for genetic analysis, so any observations of significant linkage implies a large region of the genome for further analysis, rather than yielding putative susceptibility genes straight away. Such further analysis could be futile if the original observed linkage was in fact a false positive, therefore it is considered necessary for independent replication of a locus to be observed, in order to substantiate the original finding. Such replication provides a more solid basis for the further analysis of linked genomic loci. In schizophrenia many genomic loci have been implicated by linkage analysis, but only a few of these linked regions have been consistently seen to be linked to the disorder in independent samples[117]. These regions have since been analysed for association to the disorder, with greater success than in regions where no consistent linkage was first observed.

Table 4

Linkage Guidelines	
Suggestive Linkage	expected to occur one time at random in a genome scan
Significant Linkage	expected to occur 0.05 times in a genome scan
Highly Significant Linkage	expected to occur 0.001 times in a genome scan
Confirmed Linkage	expected to occur 0.01 times in a genome scan, when performed in a subsequent independent sample

2.3.1.1.2 Association analysis

Association analysis simply measures whether or not a certain allele of a particular gene locus is found in affected individuals with significantly different frequency than in non-affected individuals over an entire sample set. Association studies primarily use SNPs as they are more common throughout the genome than microsatellite markers, yet when studied individually they suffer from low information content as bi-allelic markers. However, almost five million SNPs are to date identified and validated within the human genome, with this number expected to rise to around ten million, providing an excellent basis for haplotype construction, resulting in combinations of markers with higher information content. Haplotypes are defined as combinations of alleles from nearby DNA polymorphisms that are inherited as a block from a common ancestor. Typically, after linkage analysis has identified candidate regions, multiple SNP markers, of which it is hypothesised some will be in a degree of linkage disequilibrium (LD) with the functional mutation, are genotyped and analysed. LD means the non-random association between alleles of closely linked markers, and measures if co-occurrence of a certain sequence variation with another variation at another locus differs from chance. LD is complete when the information deduced from one variation fully correlates with the information from the other. It has been shown that significant LD between SNP pairs can be seen up to 0.1 cM, varying between populations, the longer regions of LD being observed in relatively isolated populations like those found in Finland[118]. Such isolated populations, showing wider LD intervals, would be ideal for performing association analysis[19].

The fundamental idea to utilize haplotypes on chromosomal regions identical in siblings through inheritance from the same parent (identical by descent; IBD) and LD between markers in IBD alleles in population isolates is the assumption

of a founder effect where individuals affected with the disorder share a genomic region originating from a common ancestor. Due to this, allelic heterogeneity is minimised and a smaller sample size would thus be sufficient to detect the locus. However, in complex disorders, like many psychiatric illnesses, it would be expected that the common ancestor is far more distant, and that there are a number of divergent founder alleles being inherited. This means that the shared region or regions between affected individuals would be considerably smaller and heterogeneous than those for monogenic disorders, and thus locus identification by LD remains a challenge for complex disorders.

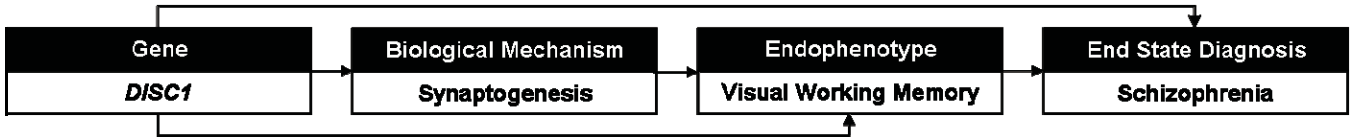
Association analysis in complex traits is greatly benefiting from the developments in the analysis programs and packages that can be used with regard to haplotypes. Until recently SNP haplotypes could not effectively be analysed for association. The individual SNPs could be placed into haplotypes by programs like GENEHUNTER[119] and SIMWALK[120], but such programs are unable to fully cope with multiple tightly linked SNPs and the LD between them[121], and are highly dependent on the quantity[122] and quality of genotype information being used. Consequently, they are best suited to calculating haplotypes for microsatellite markers in large family-based sample where both parents and offspring are genotyped. The development of genotype error checking programs that can check for Mendelian[123] and non-Mendelian[124] errors, and of haplotype analysis programs like TRANSMIT[125] and haplo-FBAT[126] has greatly advanced haplotype-based association studies. The error checking programs now mean that the raw data being analysed is of a reliable quality, which the haplotype programs can then use to estimate the haplotypes for each person even if some parental information is missing, which can then be analysed for transmission distortion with the disorder. Programs that can reliably calculate the haplotypes in population based analysis have also been developed[127] enabling testing of association using haplotypes in various study designs.

2.3.1.2 Quantitative trait analysis

Both linkage and association analysis programs traditionally use dichotomised traits which classify individuals in the analysis as either affected or unaffected, with the additional option of individuals being classified as unknown. Recently, newer programs, like Merlin[128] and QTDT[129], are starting to be able to analyse traits that are continuous variables for both linkage and association. The classic examples of such traits are height and weight, but variables can also be collected from tests of neurocognitive functions, which are deficient in individuals with schizophrenia and their relatives when compared to a control population[130]. Using such traits in the analysis of complex disorders provides a beneficial increase in the power to detect linkage or association. This is partly due to these

traits most likely being endophenotypes of the disorder in question, and therefore being representative of traits closer to the biological function controlled by genes (Figure 4, and discussed earlier). However, using quantitative variables on there own increases the power, for the simplest reason that such variables contain more information than dichotomised versions of the same traits. Yet this power can be diminished as, in family samples that have been ascertained for a high number of affected individuals, the variability of the quantitative trait in the study sample is often reduced.

Figure 4
Schematic to show the difference between testing for genetic association with an end-state diagnosis, above arrow, compared to analysing with an endophenotype, below arrow. The illustration assumes that the endophenotype would be more directly associated with the changes in biological pathways than the end-state diagnosis. As in Figure 1 the interactions are likely to be more complex, with many genes affecting a single endophenotype, which in turn can be present for more than one disorder.



2.3.2 Statistical significance

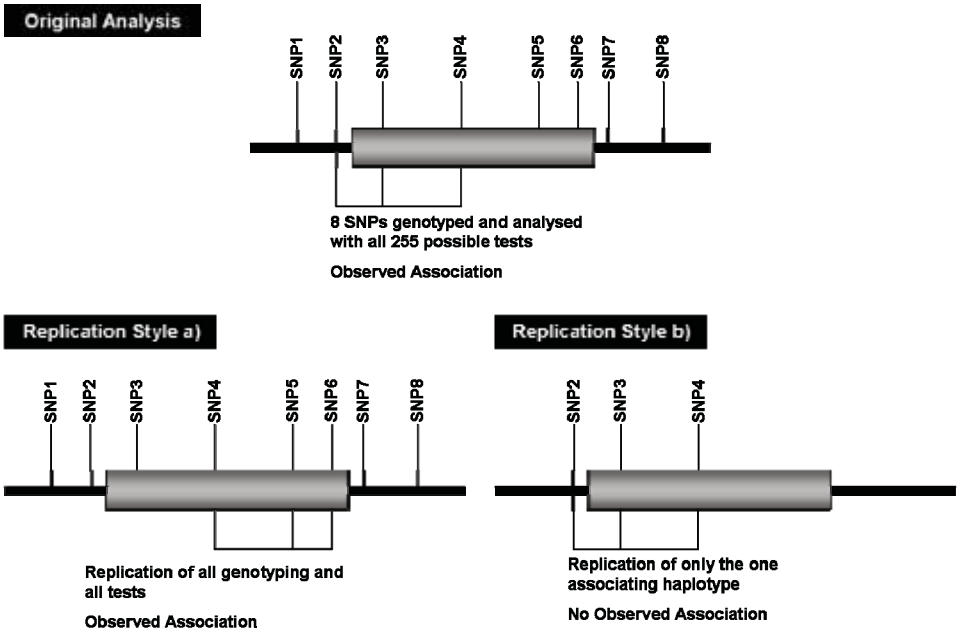
Despite the recent advances in genetic tools and analytical methods in complex disorder research, there is one main inherent disadvantage, multiple testing. In both the original linkage analysis and in association analysis many hundreds of tests are performed and with each test there is an increase in the chance of observing a type I error (false positive) which needs to be corrected for. For linkage analysis the standard way of correcting for this multiple testing is primarily to establish criteria as to what can be deemed as a significant finding from a genome-wide scan (mentioned earlier). In the case that a genome-wide scan has been performed under an extra bias that will affect the analysis, as in IV, it is common place to use simulations of the data to address the real statistical significance of the findings.

The robust standard for corrections adapted in statistical genetics is the Bonferroni correction for multiple testing, in which the observed p-value is multiplied by the number of independent tests performed. In genome-wide marker and haplotype analysis, this can mean possibly thousands of tests practically negating any finding which may be observed. However, for such haplotype analysis these corrections may be overly conservative as each test is not entirely independent as the SNPs being used are to some degree in LD[131]. Alternatively, an empiric p-value can be derived using permutation analysis for the tests performed, using the sample to simulate the analysis performed many times and thus providing an estimation of the statistical significance for the observation, by indicating how likely the observed result was to occur purely by chance. However, this is still not ideal as the empiric p-value is dependent upon the arbitrarily chosen number of permutations to be carried out by the researcher. This is illustrated from the analysis of the *G72*/*G30* genes with schizophrenia by Chumakov et al[54], where an original 20, 000 permutations provided a number of haplotypes displaying significant association, yet after a further 50, 000, 000 permutations only two haplotypes remained significant. So rather than perform such corrections and risk type II errors (false negative), it is common place to rely on independent replication in multiple study samples to verify the initial genetic findings.

This then poses a Catch-22[132] study design conundrum for future analysis, where the one method that can be used to achieve a goal, paradoxically makes that goal unobtainable. In this instance, to avoid multiple testing only the one test directly replicating the original observation should be performed. However, schizophrenia and other psychiatric disorders are complex traits and the affecting mutations and their frequencies are likely to be different between and even within populations, so even if association with the previously identified

haplotype is not found, the gene can not be excluded from the aetiology of the disorder in the replication sample. In order to perform replication for the same gene as in the previous findings, multiple SNPs and their haplotypes throughout the gene of interest should be tested. Then, even if association is found to the same gene as the previous sample, the original problem of multiple testing has not been addressed (Figure 5).

Figure 5
Illustration of the two methods that can be used to try and replicate an association finding with an allelic haplotype. Style a) replicates everything performed in the original analysis, and observes association to the trait with a different allelic haplotype, whereas style b) replicates only the positive finding from the original analysis and risks not detecting association from elsewhere in the same gene.



For association and haplotype analysis of schizophrenia the latter approach has generally been taken. Both the *NRG1* and *DTNBP1* genes have displayed an initial association to schizophrenia that has since been replicated. In most cases this replication has emerged from association data of a haplotype different or derived from the one initially presented. Therefore, if only the original haplotypes had been tested as replications in these independent samples then the

observed associations would have been missed and the support for these two candidate susceptibility genes would not be as convincing as it is today.

Therefore, researchers should continue to define the full spectrum of allelic variation at previously associating loci, yet contribute to an accumulating replication that will eventually lead to the gene being confirmed or dismissed for involvement in the aetiology of the disorder. Given guidelines would need to be based on the probability of a number of independent groups observing association at the locus using multiple tests, given that the association is a false discovery[133].

3 Aims of the study

The general aim of the study presented here has been to progress the work on the genetic risk factors that underlie the mental illness schizophrenia, using the unique nationwide study sample collected from Finland. It was intended that by focusing on a single genomic region that had multiple lines of evidence for its involvement in the disorder, the study would use a three step strategy, with each step concluding on a different aspect of the genetic aetiology of the disorder.

Validation: To determine the significance of the previously identified linkage findings for schizophrenia on chromosome 1q, through independent replication using Finnish families ascertained for schizophrenia.

Dissection: To use the candidate gene approach on a locus highlighted by significant linkage to identify genetic variations that associate with schizophrenia, and to relate any observations to the aetiology of the disorder through the use of qualitative trait components and quantitative endophenotypes.

Utilisation: To use any consistently observed associations as a conditioned factor in a genome-wide scan, in order to identify additional genomic loci that may play a role in the aetiology of schizophrenia. To be followed by the further dissection of any potentially linked region, in order to additionally discern its role in schizophrenia.

4 Materials and methods

4.1 Table of materials and methods

The details of the materials and methods used in this study can be found in the original publications according to Table 5.

Table 5

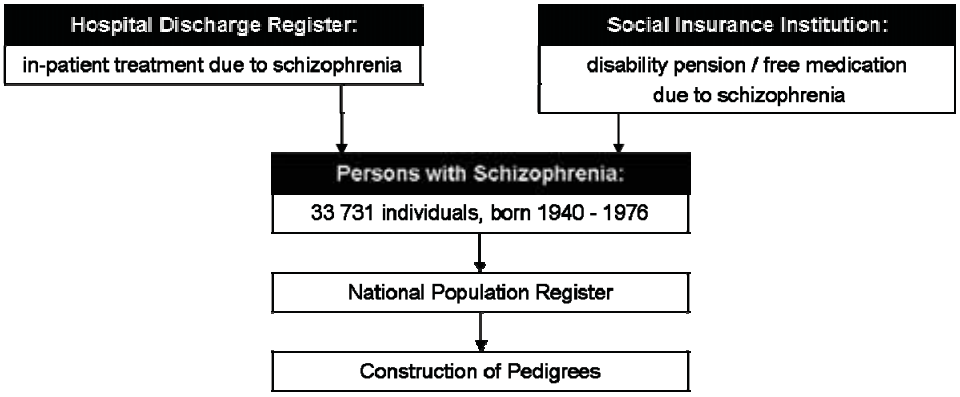
Material or Method	Original Publication
Sample Materials	
Finnish Schizophrenia Family Sample	I-IV
Linkage Replication sub-sample	I
Internal Isolate sub-sample	II
All Finland sub-sample	II
Quantitative Trait sub-sample	III-IV
HEP3 Stratified sub-samples	IV
Control sample	III-IV
Phenotype Methods	
DSM-IV Consensus Diagnosis	I-IV
Increasingly Inclusive Liability Classes	I-IV
OCCPI Trait Components	II
Neurocognitive Traits	III-IV
Transformation of Traits	III
Covariates	III-IV
Statistical Methods	
Marker and SNP Selection	I-II, IV
Genotype Correction	I-II, IV
Two-Point Linkage	I, IV
Multipoint Linkage	I
Simulation of Linkage	IV
SNP Association	I-II, IV
Haplotype Association	I-IV
Permutation of Association	I-IV
Bonferroni Correction	IV
Haplotype Construction	III-IV
Laboratory Methods	
Genotyping	I-II, IV

4.2 Study samples

4.2.1 Schizophrenia family sample

The samples used here have been collected as part of a wider project that has provided the ground work on which this study is based. Collection of this sample started in the late 1980s and in its current form is an extension of the samples previously used to perform linkage analysis for schizophrenia in Finland[7,32,88,105,134]. The sample identification and collection has remained the same throughout the years, where Finnish patients with schizophrenia born between 1940 and 1976 were identified through the hospital discharge, disability pension, and the free medication registers. Close family members of each proband where then identified through the national population register, enabling the construction of pedigrees[134,135] (Figure 6).

Figure 6
Schematic of the methods used for identifying probands affected with schizophrenia, for the collection of the study material used in all stages of this research.



The whole sample used in this study now totals 458 families consisting of 498 nuclear families that contain 2,756 individuals, of which 2,059 have been genotyped. Of these genotyped individuals, 931 are classified as affected using increasingly inclusive liability classes (LC) using criteria from the Diagnostic and Statistical Manual of Mental Disorders, forth edition (DSM-IV)[28]. LC 1 constitutes schizophrenia only, LC 2 added those individuals affected with schizoaffective disorder, LC 3 added individuals with schizophrenia spectrum disorder[136] (schizoid, schizotypal and paranoid personality disorder, schizophreniform, delusional and brief psychotic disorder, and psychosis not

otherwise specified), and LC 4 added individuals with bipolar affective disorder or major depressive disorder, with and without the presence of psychosis.
For parts of this study it was necessary to divide this large family material of 458 families into the sub-samples described here;

Linkage Replication sub-sample: Since the original publications of observed linkages on chromosome 1q in the Finnish family samples, the identification and collection of new families has continued. Therefore, there were an additional 81 families that could be used for further analysis and yet be independent of the previous observations. These 81 families consist of 501 individuals, of which 180 are classified as affected under LC 4. However, for linkage replication purposes the liability class where the best previous linkage was observed was to be used (LC 3). This resulted in 70 linkage informative families with 79 LC 3 affected sib-pairs.

Geographical sub-samples: The entire sample set could be divided into two sub-samples that are defined by the geographical origin of the family. The first sub-sample contains the families obtained from the young internal isolate population (IS) used in the genome-wide scan that showed linkage on chromosome 1q32-41[105]. This sample now consists of 179 families, made up of 1,137 individuals. The second sub-sample contains families obtained from the whole of Finland (AF), exclusive of the internal isolate, that displayed linkage on chromosome 1q42[7]. This sample now consists of 279 families, made up of 1,619 individuals. There is no overlap between these two groups and when they are analysed together they combine to create the whole sample.

Quantitative Trait sub-sample: Due to the extensive nature of the traits being measured for quantitative analysis it was only logistically feasible to test approximately half of the families from the entire sample. This sub-sample consists of 215 families containing 1,437 individuals of which 400 are classified as affected, and where 746 have undergone extensive neuropsychological assessment, including 356 offspring currently unaffected under LC 4.

HEP3 Stratified sub-samples: Schizophrenia is commonly accepted to be a polygenic disorder. Therefore in order to detect other potential susceptibility loci for the disorder the entire sample was stratified based on the HEP3 haplotype. HEP3 is a haplotype of *DISC1* identified to associate with schizophrenia through the analysis performed in stage II. The original study of 458 families was split into those families where at least one family member was predicted to have the HEP3 haplotye (n = 145) and those where no one in the family was predicted to have the HEP3 haplotype (n = 313). Only the predicted haplotypes of family members that had been genotyped were considered when making the dissection.

The two separated sample sets would then potentially be more genetically homogeneous than when they are combined, leading to the ability to further detect other prospective genetic factors influencing the aetiology of schizophrenia in this study population.

4.2.2 Control sample

In stages III and IV a sample of 60 anonymous Finnish trios, representing a random sample of the population was used in order to derive the unbiased frequencies of the HEP3 haplotype. Additionally in stage IV, this sample was combined to the whole sample for association analysis, to partially account for the ascertainment bias of the family sample.

4.3 Ascertainment of phenotypes

4.3.1 Clinical diagnosis

The diagnostic assessment used for the entire sample population is based on the analysis of all available inpatient and outpatient records for those individuals with a register based diagnosis of psychosis between 1969 and 1998, which were born between 1940 and 1976. Two psychiatrists then independently determined the consensus best estimate lifetime diagnosis, based on all available case reports, according to the DSM-IV, blind to family structure and register diagnosis. If these two psychiatrists provided conflicting diagnoses a third independent psychiatrist was used to reach the consensus. Reviewer agreement on the diagnosis has been noted to be excellent, with kappa values ranging from 95% to 99% depending on the liability class[137]. This method for determining the individual diagnoses has repeatedly been shown to be reliable[12,138,139].

4.3.2 Operational criteria checklist for psychotic illness

In addition to the consensus diagnosis made according to the DSM-IV criteria, one reviewer completed the Operational Criteria Checklist for Psychotic Illness (OCCPI). The OCCPI checklist consists of 90 items of psychopathology, pre-morbid functioning, and personal history[69,70]. Factor analysis had been performed for 30 of the 90 OCCPI items from 190 patients with schizophrenia from the IS population and 466 affected sib-pairs from the AF sample. It was found that 24 of these 30 items segregate into four factors: factor 1 “delusions and hallucinations”, factor 2 “manic”, factor 3 “negative”, and factor 4 “depressive”[137]. These factor structures were used to determine qualitative trait

phenotypes for use in stage II, where at least one of the items loading > 0.5 had to be present in the individual to count them as positive for that factor. If any of the 24 individual OCCPI items had more than two possibilities for classification, then these were also converted into qualitative traits by taking any score > 0 as meaning the individual is positive for this item.

4.3.3 Neurocognitive variables

The neuropsychological test battery, from where the quantitative neurocognitive traits were obtained, is a series of tests that uses well validated, internationally used neuropsychological instruments to evaluate an individual's cognitive ability. These tests have previously been shown to be endophenotypes for schizophrenia[66,140,141], with other endophenotypes not being collected as the tests were to be performed in the field. The test includes the Wechsler Memory Scale - revised (WMS-R)[142], the California Verbal Learning Test (CVLT)[143], and the Wechsler Adult Intelligence Scale - Revised (WAIS-R)[144]. They were administered to the subjects in a fixed order by experienced psychologists or psychiatric nurses who had received extensive training with the test battery, and all scoring was done by experienced psychologists[66]. In stage III the Visual Span forward and backward subtests of the WMS-R were used to assess visual attention and visual working memory respectively. Verbal learning and memory were assessed using the California Verbal Learning Test. From this test, using semantic clusters as a learning strategy, and the intrusive recall errors score, were logarithmically transformed to reach normality for the analyses. In stage IV only the visual working memory trait was used for analysis.

4.4 Laboratory methods

4.4.1 Sample collection and DNA extraction

For each consenting individual 20 - 30 ml of blood was drawn into EDTA tubes, from which DNA was extracted according to the standard procedure established by Blin and Stafford[145].

4.4.2 Genotyping

For the linkage analysis in stages I and IV microsatellite markers had been genotyped by ABI 377 automated DNA sequencer (Applied Biosystems). For stages I to IV SNPs were used for association analysis, these were either identified from contig alignments, from EST alignments or by sequencing for the *DISC1* locus, or from public databases for both the *DISC1* and *NDE1* loci.

Potential SNPs not located by sequencing were verified in 12 Finnish controls. In stage I and II the resulting 28 SNPs at 1q42 were genotyped in the entire sample using TaqMan (Applied Biosystems), solid-phase mini-sequencing[146] or by using array based genotyping[147]. In stages III and IV SNPs were genotyped using the Sequenom MassARRAY system[148].

4.5 Statistical methods

All microsatellite and SNP markers used in this study were checked and corrected for Mendelian errors prior to analysis using the Pedcheck program[123], and for non-Mendelian errors using the MENDEL program[124]. If such an error was observed then the genotypes for that marker in the whole family were removed.

4.5.1 Linkage analysis

The linkage analysis of each marker individually (two-point) assuming a model for genetic heterogeneity in stage I and IV was performed using the MLINK program from FASTLINK 4.1P version of the LINKAGE package[149-153]. The additional multipoint analysis of stage I being performed using the LINKMAP and HOMOG programs conditional on the marker-marker haplotype frequencies estimated with a specially written version of the ILINK program. Simulation of the linkage significance in stage IV was carried out by randomly reassigning genotypes to individuals, but keeping the genotype frequencies identical to the original analysis, to create 100 random replicates of the sample using the SIMULATE program[154,155]. Linkage analysis, as performed for the original sample, was performed on each of these replicates by using a modified version of the automated genome-wide linkage program AUTOGSCAN[156], with the derived p-values being calculated from the number of times the observed lod score was seen or exceeded in these simulations. The derivation of the p-value was performed across all scans, across all scans in each sample, and for each individual scan.

4.5.2 Association analysis

For stages I, II and IV two-point analysis was performed using the Pseudomarker program, which performs joint linkage and linkage disequilibrium (LD) analysis on a mixture of pedigrees and singletons. Pseudomarker is able to combine the power of linkage analysis with that of association, and can test for LD in general pedigrees conditional on linkage. The latter is used when it is

known that the sample already displays linkage to a particular region being analyzed for association[157]. Once two-point analysis had been performed on the SNPs the analysis of haplotypes was performed, using the TRANSMIT program. This software is able to test for transmission of a haplotype even when phase is unknown and when parental genotypes are not completely known. The TRANSMIT program is also able to compensate for the presence of linkage when using family data by the calculation of a robust variance estimate[125]. Haplotypes below a sample frequency of 3% were aggregated and counted as one haplotype when calculating the global p-value, but were ignored as being too rare as individual haplotypes. TRANSMIT performed 100, 000 bootstrap tests for all analyses, from which it derived the empirical p-values. As additional checks of the quality of the findings with the TRANSMIT program the analysis with any associating haplotypes was re-run with two conditions. First, only those families where complete genotype information was available were used in order to eliminate any false positive results based on the corrections for missing data. Secondly, the analysis was run 20 times taking randomly one affected individual per nuclear family, in order to completely eliminate any linkage affects on the association. For stage II Pseudomarker analysis was performed for all SNPs in all LC phenotypes using all three samples. However, TRANSMIT analysis was performed originally in the whole sample using only LC 4 to locate any significant haplotypes ($p < 0.05$), which were then tested in all of the phenotypes with all samples. In stage I analysis was only performed using the SNPs where linkage had previously been observed, and their flanking SNPs. For stage III and IV the TRANSMIT program was used in the same way to see how the reduced sample size from only using those families with quantitative trait information, affected the ability to detect the previously observed associations. In stage IV both Pseudomarker and TRANSMIT analysis was performed only using LC 4, with multiple testing being corrected for by using the over conservative Bonferroni method.

4.5.3 Haplotype construction

In stages III and IV it was necessary to determine the haplotypes for all genotyped individuals, yet there was no gold standard method available for performing this on data that consists of multiple tightly linked SNPs in a family based material. The most reliable way to construct these haplotypes with this data set was to use the Simwalk2 program[120], a Markov chain Monte Carlo (MCMC) and simulated annealing program for haplotype analysis. In order to increase the reliability of the predicted haplotypes, all possible markers that had been genotyped in the 1q42 and 16p13 regions were used, which included 28 SNPs and 4 microsatellite markers for 1q42, and 7 SNPs and 2 microsatellite markers for 16p13, with the entire sample of 458 families being used. The

frequencies of the resulting haplotypes were identical to those estimated from the genotypes by TRANSMIT. In stage IV the selection of which SNPs to include in the haplotype analysis was governed by testing for linkage disequilibrium between the genotyped variants. This was done by analysing the founder genotypes using the Haploview program[158], this compares each set of neighbouring markers to see how much of one marker can be defined by the other marker. Using the "solid spine of LD" criteria from this program highlighted the four SNPs that would "tag" for the whole region when analysed as a haplotype.

4.5.4 Quantitative trait analysis

In stages III and IV a direct test for transmission distortion between a haplotype and a quantitative trait was to be tested. The haplotypes were constructed using Simwalk2, as previously described, and the resulting haplotypes were recoded to form "bi-allelic markers" so as to allow the hypothesised haplotype risk allele to be tested against all other possible haplotype alleles combined. These were used within the QTDT program[129]. This is a variance component method for testing for transmission distortion of an allele with a quantitative trait. In the analysis, age, sex, and affection according to LC 4 were used as covariates, and 100, 000 permutations were performed to derive the empirical p-values.

4.6 Ethical considerations

Schizophrenia presents in signs and symptoms within the entire range of human mental activity, damaging functions regarded as specifically human. This makes the disorder highly stigmatizing to the subjects, their families, and the community. Therefore, the importance of confidentiality is primary in research into the disorder. In these studies, the principles recommended in the 1964 World Medical Association Declaration of Helsinki, and its amendments, were followed. The research was approved by the Ministry of Social Affairs and Health (Finland) and the appropriate institutional review boards, and written informed consent was obtained from all individuals. In effect this meant that all identification of the individuals who have graciously donated DNA samples to this research was through the use of numeric codes. Such a code allows for anonymous analysis by all researchers using this sample set, with only a couple of senior researchers possessing access to the password protected database containing the key to the code. Additionally, this database is only located on an internal institute network, secured by a firewall and data encryption. Furthermore, sensitive data is not allowed to be stored on laptops, portable data transfer media, or printed.

5 Results and discussion

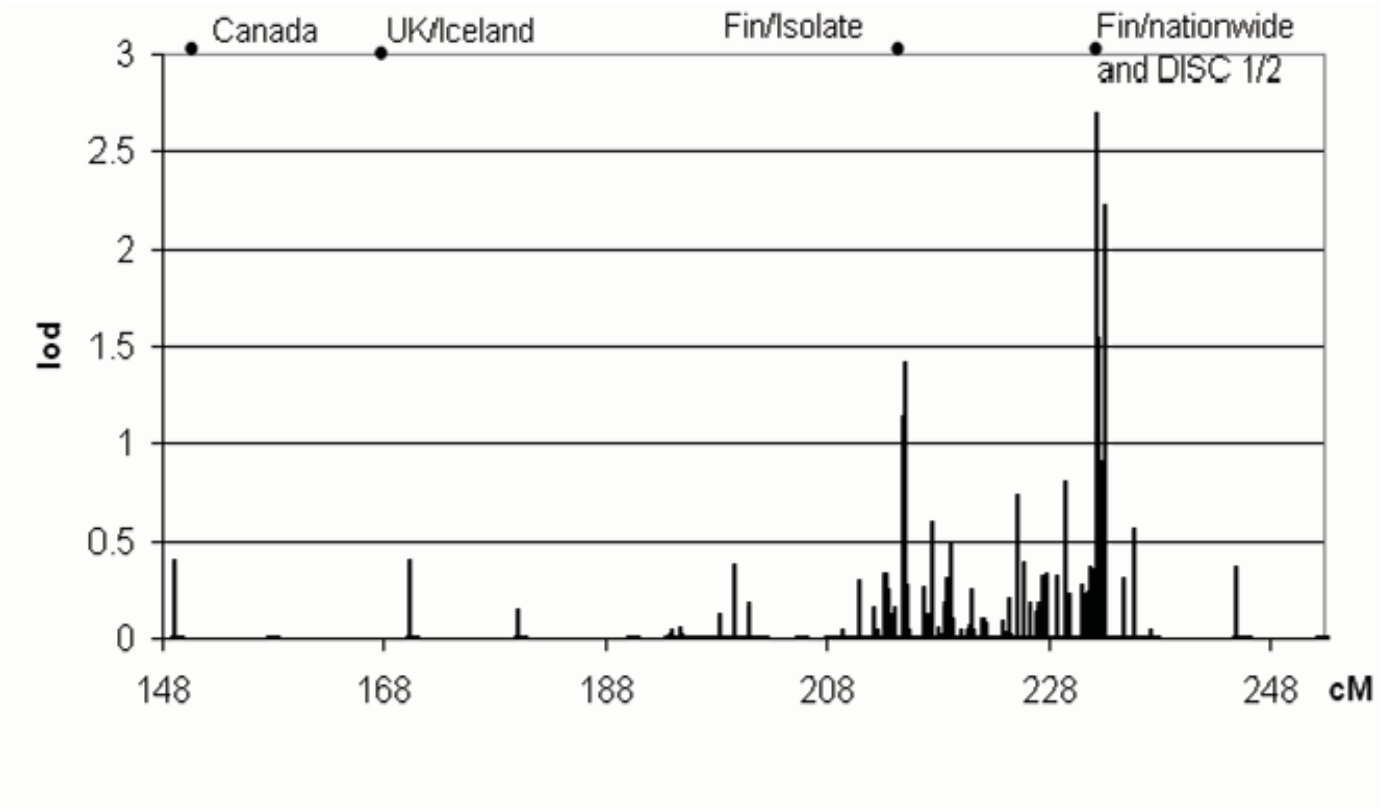
5.1 Validation of the 1q42 locus in schizophrenia

Linkage analysis of schizophrenia has highlighted a number of interesting regions on chromosome 1[3,5,8,77,80,107], two of which have been identified in samples ascertained from the Finnish population. In a study of an internal isolate population a haplotype spanning from 1q32 to 1q41 was identified[105], and markers on the 1q42 region were observed to be linked to schizophrenia in the study sample of 168 Finnish families exclusive of this internal isolate[7].

Since these studies the sample material for the analysis of schizophrenia in Finland has been expanded, with the addition of 70 families containing a total of 79 affected sib pairs (ASPs). Initially this new independent sample was to be used to study chromosome 1 in order to verify through replication the genomic regions that had been previously linked to the disorder. A dense set of 300 polymorphic markers were analysed including all the microsatellites that had been previously used in the two prior analyses of schizophrenia in Finland[7,105], with the density of markers being concentrated around these two Finnish findings at 1q32 and 1q42. Two-point and multipoint linkage analysis was performed for all markers, and as some of the markers were SNPs, two-point and haplotype association analysis was performed for those providing evidence of linkage.

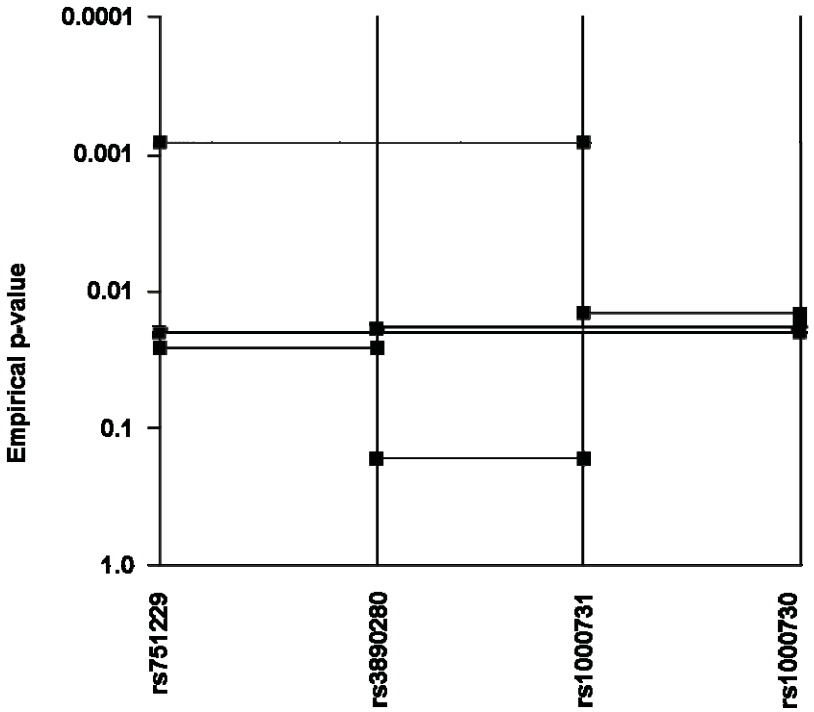
Linkage was observed at the 1q42 locus (Figure 7) using a dominant affected only model for LC 3. The best two point lod score (lod = 2.70) was obtained for a SNP located within intron 9 of the *DISC1* gene (rs1000731). A neighbouring SNP 1.5 kb towards the centromere (rs3890280) also provided some evidence for linkage (lod = 2.30) (I). The SNP rs1000731 is located only 67 kb from the microsatellite marker D1S2709 that provided the strongest evidence for linkage in the previous Finnish finding on 1q42[7].

Figure 7
Graph of two-point linkage findings on chromosome 1q, showing linkage observations with respect to the previously published findings on this chromosome.



Two-, three- and four-marker haplotype association analysis was performed for the two most significantly linked markers and their two flanking SNPs (Figure 8). Two-point association analysis, conditional on the presence of linkage, displayed little evidence of association, (rs751229, p = 0.022; rs3890280, p = 0.20; rs1000731, p = 0.090). Additionally, the two-marker haplotype containing only the two most strongly linked SNPs did not provide evidence of association (p = 0.31). However, a three marker haplotype containing the SNP rs751229 in addition to these two SNPs provided significant association (p = 0.00090), being over-transmitted to affected individuals (I).

Figure 8
Graph to show the empirical p-values from the association test using haplotypes of four SNPs located in intron 9 of *DISC1* on 1q42. Style adapted from Schwab et al[23]



For the 1q42 genomic region, the most likely candidate genes are *DISC1* and *2*, with the best evidence from three independent studies being intragenic of these genes[3,7](I), with a further two showing linkage in the promoter region of *DISC1*[5,8]. This would appear to contradict the evidence from simulation studies

demonstrating poor resolution of linkage peaks in the efforts to position a complex disease gene[134,159]. However, although weak evidence of association was observed for markers under this linkage peak, it cannot be excluded that nearby susceptibility variants are partly or wholly responsible for the observed linkage signals.

The microsatellite marker D1S2709 that provided the original observation of linkage at 1q42 in the Finnish population failed to show significance in this new study sample (two-point lod = 1.00) (I). So, although the study sample was from the same Finnish population, replication would not have been obtained using the same set of microsatellite markers, and it was only by using a denser marker map and by including SNPs, that evidence for linkage was observed that clearly exceeded the genome-wide replication threshold (point-wise p of 0.01 ~lod 1.18) suggested by Lander and Kruglyak[116]. This underlines difficulties involved in interpretation of linkage findings in complex traits, and brings additional queries to the validity of a previous attempt to replicate linkage on chromosome 1q using only 16 microsatellite markers[109]. As here linkage was only detectable using a very dense marker set, and would not have been observed if only the 16 markers used in the multi-centre study would have been used.

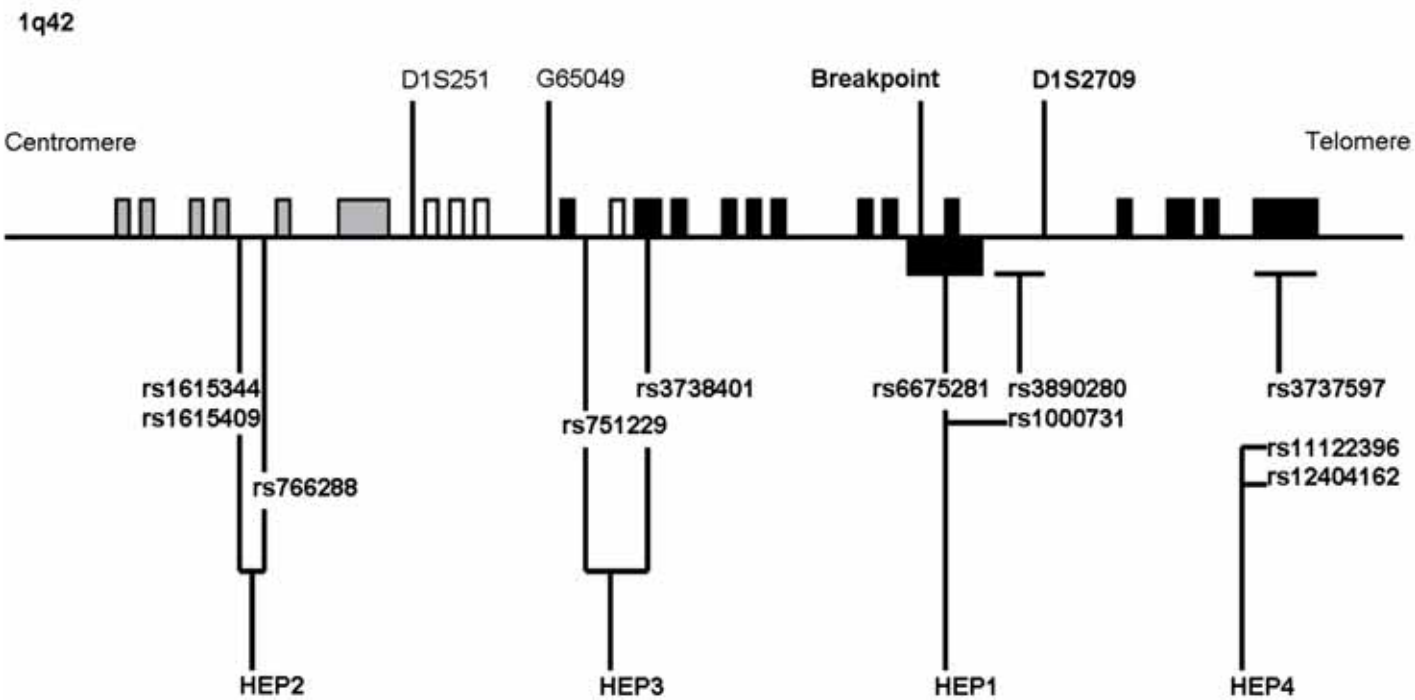
5.2 Dissection of the 1q42 locus in schizophrenia

To further analyse the 1q42 locus, 28 SNPs located within 600 kb were monitored for two-point and haplotype association to schizophrenia, in the entire sample material available. Two-point analysis showed that one SNP (rs3737597), with a minor allele frequency of 2%, located in *DISC1* exon 13 displayed consistent suggestive association with schizophrenia, displaying a p-value of 0.0021 (Chi square = 9.44, 1 df) for LC 2, but also displaying p-values < 0.01 for the other liability classes (II).

Haplotype analysis was performed for the whole region using the broadest phenotype (LC 4), in order to include all possible phase information and statistical power of the sample, and used the sliding window technique of systematically taking two to three neighbouring SNPs in turn for the analysis. One haplotype (HEP1) was observed to be significantly over-transmitted (p = 0.00045) to affected individuals, and is the haplotype identified in the replication of the original linkage finding (I). Three other haplotypes displayed significant under-transmission to affected individuals. The first (HEP2, p < 0.000010) was located within the *TRAX* gene, while the other two (HEP3, p = 0.0031; HEP4, p = 0.00057) were both located in *DISC1* (Figure 9) (II). Further, the HEP3 haplotype displayed significant evidence for under-transmission only to affected females (p = 0.00024) compared with under-transmission to affected males (p =

0.38). While, for HEP4 the TRANSMIT program predicted that these SNPs would only produce two haplotypes. Consequently, the alternative haplotype was over-transmitted with a significant p-value of 0.00057, but was ignored by the TRANSMIT program due to a sample frequency below 3%. Therefore, within the six SNPs located in *DISC1* exon 13, associations were observed for a single SNP (rs3737597) and a number of 2 SNP haplotypes (II), but as all associations seen in this region involve rare polymorphisms (sample frequency < 3%) conclusions should be drawn with caution.

Figure 9
Schematic diagram to show the relative locations of the two-point association and the four identified haplotypes on chromosome 1q42. The exonic structure of *TRAX* (grey boxes), and the *DISC* genes (black boxes) are also shown in relationship to the intergenic exons (white boxes).



Two validation tests for these findings were performed using the three common haplotypes (HEP1, HEP2, and HEP3), to ensure that these findings are true, and do not represent linkage or errors caused by estimation of the missing data. Initially the transmission distortion of these haplotypes was analysed in only those families where both parents were genotyped (n = 147 nuclear families). Only HEP3 continued to display the statistical significance already observed (p = 0.00084), but although evidence for sex differences in transmission distortion remained, this difference was not as significant as in the complete study sample of 458 families (Female, p = 0.0020; Male, p = 0.018). Tests randomly taking only one affected offspring per family were additionally performed. This was carried out 20 times for each haplotype taking the mean value of these as the average. Only HEP1 displayed no significance (p = 0.17), while both HEP2 (p = 0.0040) and HEP3 (p = 0.035) still showed some association to the LC 4 phenotype, with HEP3 still displaying sex dependent differences in transmission (Female, p = 0.0065; Male, p = 0.38) (II).

Throughout these analyses, the HEP3 haplotype appeared to represent the most robust finding. This haplotype spans 62 kb from intron 1 to exon 2 of *DISC1*, highlighting the gene as contributing to the aetiology of schizophrenia. With the observed sex differences in the association putatively implying that this genetic region could contribute to the sex differences found in the schizophrenia phenotype[36,45,46,160]. Mechanisms for sex differences, such as imprinting[161] and hormonal effects[162], can biologically present in gene expression.

The haplotype HEP2 is located within intron 4 of *TRAX* and displays levels of significance similar to that of the HEP3 finding, and in the same pattern for all tests except one. When HEP2 was tested in only those families with complete genotype information no significant association was observed. Despite this the other association findings with HEP2 indicate that it is still worth investigating further. In contrast, when the HEP1 haplotype was analyzed for association using complete families or randomly selected trios no further evidence was observed, suggesting that this finding may be due to the previously observed linkage findings in this localised region.

To extend the analysis of the role of these three putatively associated haplotypes they were tested further by analyzing trait components as well as the end state diagnosis, using the remaining three liability classes, the four factor groups, and the 24 OCCPI items. HEP1 was found not to show further associations to any other phenotype or sub-sample. While, in the analysis of the sub-samples, both HEP2 and HEP3 were significantly associated only within the AF population, with no suggestive results being seen in the IS sample (II). This is consistent with previous findings that have observed linkage with the AF population to

1q42, while the IS population displays linkage to 1q32-41[7,105]. Analysis of HEP2 and female specific HEP3 using the component traits, showed that these haplotypes mainly associate to traits representing delusions, hallucinations, and negative symptoms (II). Yet, as trait-components are highly dependent upon the end-state diagnosis, then it is possible that the associations observed with these traits are really a reflection of the association already observed with schizophrenia. However, it is well characterized that more severe negative symptoms in individuals with schizophrenia correlates with poorer cognitive ability[163]. Therefore in order to take analysis of the 1q42 locus further, the two most consistent haplotypes (HEP2 and HEP3) were tested with five traits of neurocognitive ability that had previously been shown to be linked to the *TRAX/DISC* neighbouring region 1q41[10,106].

Endophenotype analysis was carried out in a sub-sample of the original analysis in which 746 individuals were assessed with a neuropsychological test battery. To evaluate the effect of the reduced sample size on the ability to detect the original association with these haplotypes, analysis using the same program and specifications from the original finding (II) was performed, within this extensively phenotyped sample. Despite the reduction in sample size the association signal, in relation to schizophrenia diagnosis for both HEP2 and HEP3 remained at the suggestive level ($p = 0.0082$ and 0.072 respectively) (III).

Next, association analysis of the HEP2 and HEP3 haplotypes was carried out using the quantitative neuropsychological test variables: visual working memory, visual attention, verbal learning and memory, using semantic clusters as a learning strategy, and intrusive recall errors. It can be hypothesized that the use of more informative quantitative traits, if truly associated with clinical vulnerability and the studied gene variations, has the potential to display stronger association. The haplotype HEP3 displayed up to 50-fold more significant association to the traits of visual attention and working memory, with p-values 0.0079 and 0.0013 respectively, than to schizophrenia, but did not show association to any of the three traits of verbal learning and memory (III). The haplotype HEP2 showed no association to any of the five traits tested. In order to determine whether these associations are primarily to the trait itself or rather reflects the effect of schizophrenia on the trait distribution, the association between HEP3 and the two traits of visual memory functions was analysed separately for affected and unaffected individuals. Only for visual working memory did both affected and unaffected individuals contribute to the observed association in the combined study sample, p-values 0.034 and 0.048 respectively (Table 6). Furthermore, in analysis with affection status HEP3 displayed sex differences in its transmission distortion, only being significantly under-transmitted to affected females. In order to investigate if similar sex dependent

effects were present in the association to these quantitative traits the analysis was repeated, once only including male offspring and then only including female offspring. The association to the visual working memory trait was only significant in males (males, $p = 0.0060$; females, $p > 0.10$) (Table 6) (III).

Table 6
Empirical p-value results from analysis of the two haplotypes and LC 4 or the two quantitative traits of visual memory functions used in this study.

	HEP3	HEP3 [a]	HEP3 [b]	HEP3 [c]	HEP3 [d]
Whole Sample:					
Clinical diagnosis	0.0031	0.38	0.00024	NA	NA
Quantitative trait sample:					
Clinical diagnosis	0.072	0.34	0.096	NA	NA
Visual attention	0.0079	0.025	0.019	0.0025	> 0.10
Visual working memory	0.0013	0.0060	> 0.10	0.034	0.047

[a] Analysis using only male offspring
[b] Analysis using only female offspring
[c] Analysis using only offspring affected under LC 4
[d] Analysis using only offspring currently unaffected under LC 4

Although the results of this study support the hypothesis that *DISC1* associates with visual working memory and attention, the finding that HEP3 associates with poorer performance appears counter intuitive. The HEP3 haplotype was observed to be under-transmitted to affected individuals. Unaffected individuals, when compared to individuals with schizophrenia, have better performance in both visual attention and visual working memory. Yet the HEP3 haplotype does not associate with superior but rather with poorer performance. In addition there is the complication that the observed associations each have sex dependence, being only to females when analysing the end state diagnosis, and only to males when analysing visual working memory.

It was therefore necessary to re-evaluate the original finding of association to schizophrenia. To this end, the two SNPs that constitute the HEP3 haplotype (rs751229, rs3738401) were genotyped in a group of 60 control trios, representing anonymous samples from the Finnish population. From this sample set the HEP3 haplotype and transmission frequencies were determined, which were lower than in the schizophrenia family sample (Table 7). Using a Chi-square test, the relationship between the observed number of transmissions to affected individuals in the whole sample and the expected number were re-examined based on the transmission frequencies from the control trio sample (Table 8). Although this analysis has less power for the detection of association it would be indicative of the type of results expected to be observed by analysis similar to the one performed by the higher power TRANSMIT program, but using allele frequencies determined in a control population. These results suggest that due to HEP3 having a higher frequency in the schizophrenia sample than controls, the original number of expected transmissions was upwardly biased, which in turn led to the epiphenomenon of the observed under-transmission to affected females. In the light of the present results a more parsimonious interpretation of the data would be that the HEP3 haplotype actually confers risk to males in this study population, and that this risk affects schizophrenia and its related disorders through a role in visual working memory.

Table 7
Haplotype and observed transmission frequencies of the HEP3 haplotype in a control sample and the whole sample.

Sample	Haplotype Frequency	Transmission Frequency
Control Trios	0.070	0.072
Whole Sample	0.088	0.075
Male Offspring	0.097	0.080
Female Offspring	0.077	0.066

Table 8
Re-examination of the difference between the observed and expected number of transmissions to affected offspring in the whole sample. a) The original observation. b) Using an expected transmission frequency based on that of a control sample.

a)

	Whole Sample		Male Offspring		Female Offspring	
	Observed	Expected	Observed	Expected	Observed	Expected
HEP3	125	145	81	85	44	60
p-value		0.086		0.61		0.039

b)

	Whole Sample		Male Offspring		Female Offspring	
	Observed	Expected	Observed	Expected	Observed	Expected
HEP3	125	109	81	66	44	43
p-value		0.12		0.050		0.92

In the whole sample there were 1678 transmissions to affected offspring, of which 1006 were to males and 672 to females.

5.3 Utilisation of the 1q42 locus to identify additional loci in schizophrenia

It is generally accepted that schizophrenia is a polygenic disorder, and the concept that identification of the first candidate gene would greatly facilitate the identification of others has been proposed. The genome-wide scan data from the whole sample was analyzed ascertained for the presence of the *DISC1* HEP3 haplotype in each family. Although from the earlier findings (II, III) it would be possible to hypothesise about the causative mechanism represented by HEP3, a more conservative criterion for dividing the sample population was used. The original study of 458 families was split into those families where at least one family member was predicted to have the HEP3 allelic haplotye (n = 145 families) and those where no one in the family was predicted to have the HEP3 allelic haplotype (n = 313 families). It can be hypothesized that since locus heterogeneity certainly exists between families this ascertainment strategy could provide some added power to identify genes involved in the molecular pathogenesis of schizophrenia.

Genome-wide data from 443 microsatellite markers was analysed in the resulting two samples and the whole data set. With these analyses being performed in all four increasingly inclusive liability classes, and using both dominant and recessive models. This meant that a total of 24 genome scans were being tested and therefore, simulation was essential in order to derive the significance of any findings in the face of such multiple testing, and to account for the additional biasing caused by the conditioned separation. This was performed by simulating the genotypes for the whole sample 100 times, and then performing all analyses for the separate samples and models for each replicate. This simulation was also used to show that the sample sets linkage information content do not greatly differ, ensuring that one sample was not liable to provide over-inflated lod scores of no significance (Figure 10).

Most of the study sample represented families that were used for the previous published Finnish genome-wide scans for schizophrenia[7,32,88,105], but also contained new families that have been collected as an extension of those studies (I). When the complete study sample (n = 458) was used in the linkage analysis across the eight models tested, evidence for linkage (lod > 3) emerged for two genomic loci: D1S2709 (lod = 3.64) located at 1q42 intragenic of the *DISC1* gene, and D5S647 on 5q12.3 (IV). The 1q42 locus was the region where linkage had previously been observed in the Finnish family study sample[7](I), and here the most significant linkage was also displayed under the dominant model of LC 3 as was seen in these original analyses. The marker D5S647 is located 71 cM from the previously observed linkage on 5q33 in the Finnish population[88], but is

within the chromosome 5q11-q13 region originally implicated in Icelandic, UK and Canadian study samples in 1988[13,14].

Once the sample had been conditioned for the HEP3 haplotype, three loci provided evidence for linkage (lod > 3): 1q42 (D1S2709; lod = 3.31), 10q21 (GATA101E02; lod = 3.58) and 16p13 (D16S764; lod = 3.17) (Table 9), all three within the sub-sample positive for the HEP3 haplotype (n = 145 families) (IV). It was expected that D1S2709 would display a high lod score since this marker showed the initial significant linkage in Finnish families for 1q42 that led to the discovery of the HEP3 haplotype, on which the samples used here were stratified. On chromosome 16p13 two neighbouring markers displayed lod > 2 for the dominant model under LC 1, with one of them showing lod > 3. Most interestingly, this region of the genome contains a gene encoding a known *DISC1* binding protein (*NDE1*)[164-166]. Furthermore, linkage to this locus was observed in Finnish bipolar disorder families[167], implying the relevance of the *DISC1* "pathway", not just *DISC1*, would be important in the aetiology of schizophrenia, and other mental disorders. The GATA101E02 marker on chromosome 10 displayed its evidence of linkage under the dominant model with LC 4, and is located at 10q21 in virtually the same region previously found to display significant linkage to schizophrenia in Ashkenazi Jews[50] and suggested in other studies[48,49].
No other locus displayed linkage lod > 3 after the HEP3 conditioning, however, nine further loci displayed linkage lod > 2. Markers on 1p13, 2q32, 6p21, 7q22, 8p22, 11q22 and 12q21 were suggested in those families with the HEP3 allele, and markers on 5q12 and 8q24 were observed in the families negative for HEP3 (Table 9) (IV). Of all these 12 loci, 8 are located at or around 20 cM from genomic regions previously identified in linkage studies of schizophrenia (Table 9), and include the loci for the *DTNBP1* (6p21)[24], *NRG1* (8p22)[42], *GRM3*[34] and *RELN*[31] (7q22) (Figure 11) genes that have all been associated to schizophrenia (for review see ref[1]).

Table 9
Significance of the lod scores > 2 observed in the two stratified sub-samples, derived from the simulation of the analyses performed. Additionally showing the locations of the nearest previous schizophrenia linkage findings where appropriate.

Location	lod score	p-value	Sub-Sample	Previous	Candidate Genes	cM to Previous
1p13.2	2.20	0.67	HEP +	1p13.3-q23.3 [a]		0
1q42	3.31	0.060	HEP +	1q42 [b]	*DISC1*	0
2q32.2	2.06	0.77	HEP +	2q22/2q37 [c]		47.7/41.9
5q12.3	2.38	0.53	HEP -	5q11.2-q13.3 [d]		0
6p21.2	2.16	0.67	HEP +	6p23/6p22.3 [e]	*DTNBP1* [k]	22.2
7q22	2.25	0.61	HEP +	7q22 [f]	*GRM3/RELN* [f]	0
8p22	2.00	0.82	HEP +	8p22-p11 [g]	*NRG1* [l]	0
8q24.13-q24.21	2.58	0.42	HEP -	-		-
10q21.3	3.58	0.020	HEP +	10q22-q24 [h]		15.4
11q22.3	2.80	0.28	HEP +	11q14-q21 [i]		17.5
12q21.1	2.19	0.67	HEP +	12q24 [j]	*DAO* [j]	33.9
16p13.11-p12.1	3.17	0.080	HEP +	-	*NDE1*	-

a2 b3-9 c2,10 d13,14 e2,22-27 f31-34 g40-44 h48-51 i3,49 j 54 k23,24,27,60-64 l42-44,67,68

Figure 10
Histogram to show the linkage information content of the three samples used here, showing the significance expected for a range of lod scores, given the number of tests performed. The thick solid black line represents the significance across 24 models performed. The thick dashed line represents the significance across the 8 models performed on the whole sample. The thin solid line represents the significance across the 8 models performed on the sample of families who did not carry the HEP3 haplotype, and the thin dashed line represents the significance across the 8 models performed on the sample of families with the HEP3 haplotype.

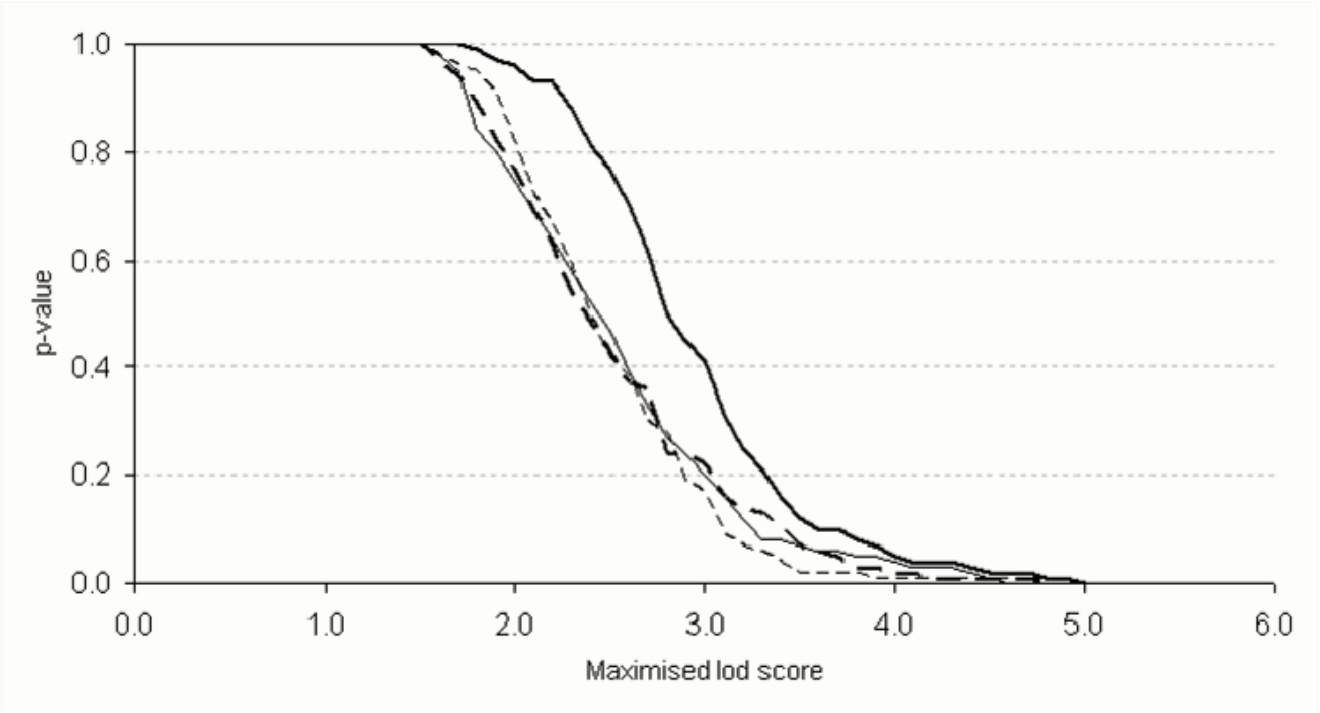
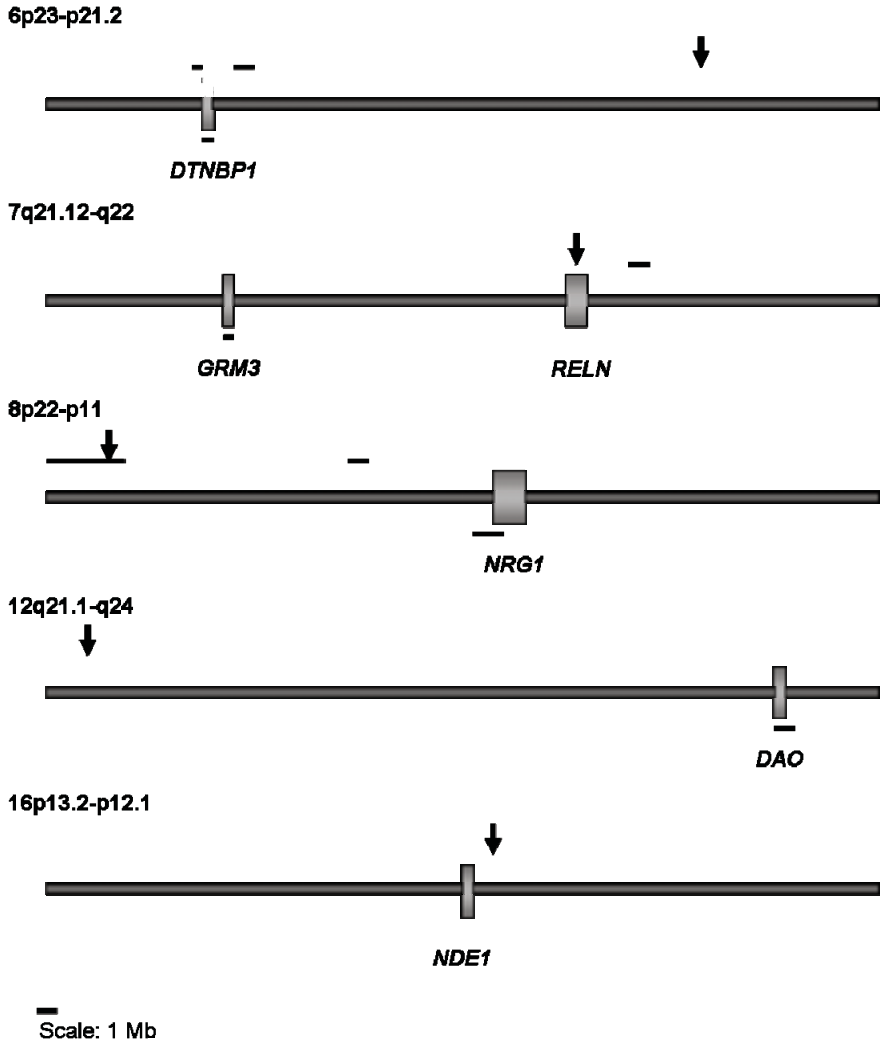


Figure 11
Schematic diagram to highlight the primary regions of interest indicated by performing a genome-wide scan conditional on a *DISC1* haplotype. Locations of the microsatellite markers that provide the evidence of linkage are indicated with downward pointing arrows. Thick black lines above each chromosome represent previous linkage findings and thick black lines below represent previous association findings. All schematics show 40 Mb, with relevant genes illustrated for each chromosome.



By controlling for *DISC1* status in a large nationwide collection of Finnish schizophrenia families it was possible to detect linkage to several loci containing previously identified candidate genes, not observed in the whole sample. This finding would provide some support to the hypothesis that several of the associated candidate genes reported actually work in concert. Therefore, such an approach could guide the gene identification efforts in other study samples and hopefully facilitate formulation of a comprehensive model of the complex genetic background of schizophrenia.

However, the observed pattern of linkage raises an important question about the linkage results presented here. In all three sample sets analyzed, the average number of individuals per family was approximately equal (n = 6), as was the average number of affected individuals per family (n = 2). Additionally, it was shown through simulation that all three sample sets were of approximately equal information content for linkage (Figure 10). Therefore, an equal number of lod scores > 3 and lod scores > 2 in each of the three samples should have been observed. Yet this was not the case as we noted a total of 10 loci in the families carrying the HEP3 haplotype, and only 2 loci in the families which did not carry the haplotype. With 6 of these loci also being observed when these two samples are combined into the whole sample. This would suggest that there is something in the stratification that underlies the bias towards observing linkage in one sample and not the other. With that strategy ultimately representing the role of *DISC1* in the aetiology of schizophrenia, a number of potential hypotheses can be implicated. Initially, it would suggest that the sample families where HEP3 has been identified as a risk factor could have a higher genetic liability to this complex disorder for which they are potentially more homogeneous, and that in those families where other genetic factors act without the *DISC1* risk allele, a greater number of environmental risks may be acting. However, such a hypothesis based on this observation may remain theoretical until more is determined about the causes of schizophrenia.

The linkage on 16p13 immediately implicated the potential involvement of the *NDE1* gene, since the corresponding polypeptide has been shown to bind to the *DISC1* protein[164-166]. Seven SNPs over the 75 kb of the *NDE1* gene were genotyped and analyzed along with their corresponding haplotype consisting of four SNPs that "tag" for this gene. Only 2 SNPs (Table 10) and the tag-haplotype (Table 11) displayed association < 0.05, although neither the SNPs nor the haplotype were significant after Bonferroni correction for the 8 tests performed at this stage (IV). As the observation of association to *DISC1* had displayed sex dependent effects it was tested if such effects may also affect the association with the *NDE1* gene, since it is encoding a protein with an established biological connection with *DISC1*. Such an effect was observed,

with all SNPs (Table 10) and the tag-haplotype (Table 11) displaying p-values < 0.05 only for female offspring. After correction for multiple testing the association observed with the tag-haplotype (p = 0.00046) remained significant (p = 0.011) for the total 24 tests performed (IV). The risk allele of the tag-haplotype comprises the CGCC alleles of the SNPs rs4781678, rs2242549, rs881803, and rs2075512 respectively, and is present in the families ascertained for schizophrenia at the frequency of 30% while in the control trio sample from Finland it has a frequency of 19%.

Further, as *DISC1* had been shown to be associated with a neurocognitive measure of visual working memory in this sample, it would logically follow that an interacting gene may also associate to the same trait. In the analysis of *DISC1*, it was hypothesized that association to the quantitative variable, if it is truly associated with clinical vulnerability and the studied gene variation should be stronger than that to the end-state diagnosis. Such an increase was not seen for the *NDE1* tag-haplotype, yet the analysis of only female offspring approached the 0.05 level of significance (Table 11) (IV).

Table 10
Observed empirical p-values for the seven analyzed SNPs located within the *NDE1* gene located on chromosome 16p13. Association was tested between the genotype and the end-state diagnosis of the broadest liability class LC 4, for the whole sample and then separated depending on the gender of the offspring.

	rs4781678	rs6498567	rs2242549	rs881803	rs1050162	rs2075512	rs11130
Whole Sample	0.051	0.0086	0.27	0.021	0.17	0.078	0.34
Males	0.66	0.48	0.40	0.53	1.00	0.53	0.44
Females	0.0052	0.0024	0.018	0.0013	0.014	0.0044	0.023

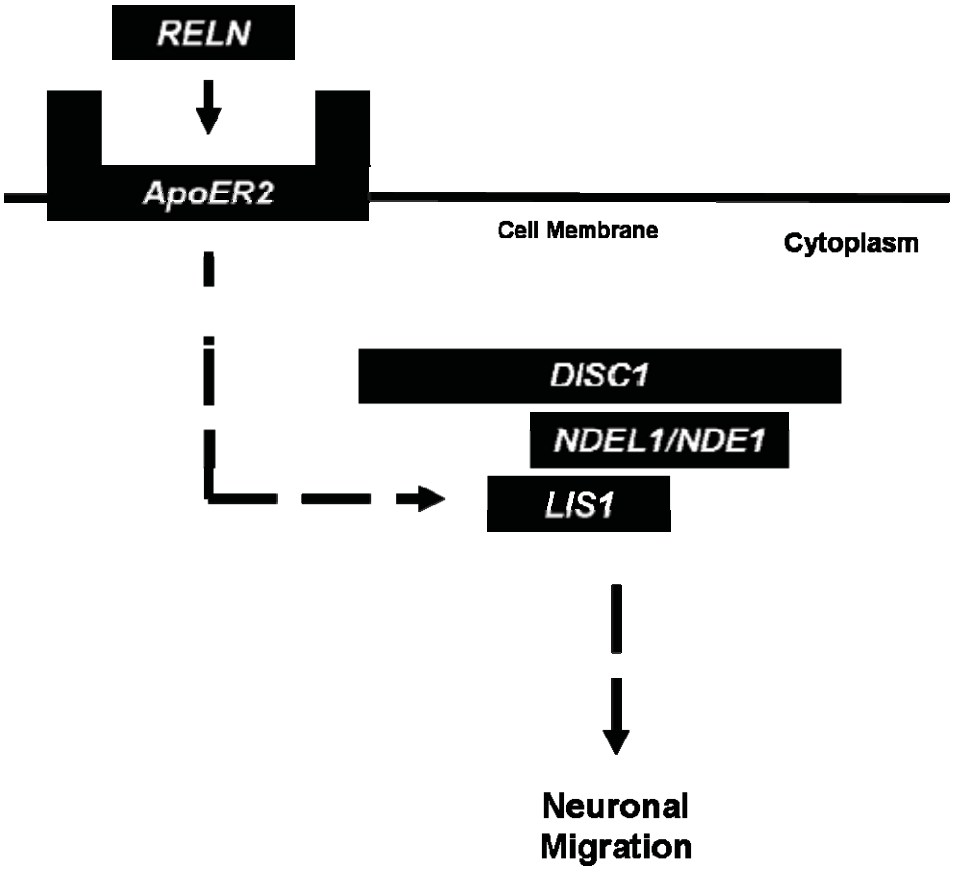
Table 11
Empirical p-values from the analysis of the *NDE1* tag-haplotype with end-state diagnosis and the neurocognitive variable representing visual working memory.

	tag	tag [a]	tag [b]	tag [c]	tag [d]
Whole Sample:					
Clinical diagnosis	0.0065	0.93	0.00046	NA	NA
Quantitative trait sample:					
Clinical diagnosis	0.046	0.25	0.016	NA	NA
Visual working memory	0.083	> 0.10	0.055	> 0.10	0.068

[a] Analysis using only male offspring
[b] Analysis using only female offspring
[c] Analysis using only offspring affected under LC 4
[d] Analysis using only offspring currently unaffected under LC 4

These findings would implicate that two interactive proteins, *DISC1* and *NDE1*, would work in concert in the genetic aetiology of schizophrenia and, potentially mediate some of this effect via deficits in working memory or other cognitive functions[168,169]. The currently available functional information for both *DISC1* and *NDE1* lend this hypothesis further credence and would expand the hypothesis of a "defective" pathway in functions relevant for schizophrenia. Not only do the proteins of these two genes interact[164-166] but they also interact to form a complex of proteins with *LIS1*[165], with *NDE1* appearing to be interchangeable with its homolog *NDEL1*[164,166]. It is known that a *LIS1/NDEL1* complex functions in neuronal migration regulated by signalling from another schizophrenia candidate gene, *RELN* (Figure 12)[170]. Further, mouse models with differing *LIS1* heterozygous mutations have been shown to have cortical and hippocampal disorganization, and have impaired spatial learning and coordination[171,172], while mouse models with *NDE1* homozygous mutations have been shown to have less neurons in the cerebral cortex, with thin cortical layering[173]. Both cortex and hippocampus are highly implicated in schizophrenia[174,175], with *DISC1* having been shown to be additionally associated to reductions in grey matter in these two regions[82,176]. This convergence of multiple lines of evidence starts to implicate not just *DISC1* but a "*DISC1* pathway" that also incorporates *NDE1*, in the aetiology of schizophrenia, potentially through underlying deficits in learning and memory.

Figure 12
Illustration of the relationship between *RELN* signaling and the potential functions of *DISC1* and *NDE1* in neuronal migration with *LIS1*. Dashed arrows represent intermediate stages between the process of *RELN* signaling and its ultimate response.



5.4 General discussion

In the beginning of this study there was only a balanced translocation that had been identified in a large Scottish family with major mental illness, which implicated not just 1q42 but also 11q14.3. The initial observation of significant linkage in the Finnish population concentrated the research effort to the 1q42 locus. Observations of linkage for 1q42 to schizophrenia have since been observed in the Taiwanese[5], and to bipolar affective disorder in a UK and

Icelandic sample[8], and furthered with the addition of independent replication in a second Finnish sample (I). Linkage to the neighbouring region of 1q41 was also noted in two independent Finnish samples when using quantitative neurocognitive traits that represent cognitive ability[10,106], endophenotypic traits that are deficient in individuals with schizophrenia and their relatives.

Located at the 1q42 locus are two novel genes that were initially identified in 2000. They are located on opposing strands of the genomic sequence, with both of them being directly disrupted by the translocation in the Scottish family. This led to their nomenclature assignment of Disrupted in Schizophrenia 1 and 2 (*DISC1* and *DISC2*). Due to this, and the localisation of the observed linkage peaks in the Finnish family material being intragenic of *DISC1*, they became the prime candidate genes for further analysis with regard to schizophrenia. In addition, the Translin-Associated Factor X (*TRAX*) is located centromeric and in the same orientation as *DISC1* and was identified as a further potential candidate gene when intergenic splicing was found to form fusion transcripts between *TRAX*, *DISC1*, and combinations of four intragenic exons located between these two genes[177]. It has also been shown that orthologs of *DISC1* are highly conserved in genomic structure and in its location close to the *TRAX* orthologs on the *Mus musculus* (mouse)[178] and *Takifugu rubripes* (pufferfish)[179] genomes, implying some functional significance for the physical vicinity of the *TRAX* and *DISC1* genes.

Dissection of the biological role of these genes has greatly furthered their positions as candidates for a role in the aetiology of schizophrenia. Protein-protein interaction studies show that *DISC1* acts as a "scaffold" for proteins known to be involved in numerous neuronal functions, impacting on neurite outgrowth, neuronal migration, cytoskeletal modulation, and signal transduction[164-166,180-183]. The *DISC2* gene is not known to encode a protein but is thought to act through its RNA as a regulator of *DISC1*[107]. The *TRAX* protein is known to form a brain enriched complex with Translin, that can bind single stranded DNA and RNA through which it is involved in protein regulation[184] and, consequentially, in development and function of the nervous system.

Since the observations in stage II of *TRAX*/*DISC* allelic haplotypes associating with schizophrenia, a number of independent replications have been published. In each case differing variants have been observed to be associated with the disorder in question, which includes schizophrenia[4,6,9,82], schizoaffective disorder[4,6], and bipolar affective disorder[6,9]. However, two studies in the Japanese population have failed to find evidence to support a role for *TRAX*/*DISC* in the aetiology of schizophrenia[185,186], although one of these studies only evaluated promoter variants of *DISC1* rather than the full locus profile[186]. During the dissecting of the association findings with quantitative endophenotypes in stage III, an independent

concurrent study in a sample of Finnish twins also analysed the HEP haplotypes for association to a wider selection of traits derived from the same neuropsychological instruments as used in the family data and from magnetic resonance imaging (MRI) data. This study also observed association for these haplotypes with schizophrenia and visual working memory, and observed association to decreased grey matter volume in the prefrontal cortex[176]. It had previously been observed that a *DISC1* SNP variant and its haplotypes associate with decreased grey matter volume in the hippocampus[82], these two brain regions are those most often associated with schizophrenia[174,175,187,188]. Additionally, at the time of publication of study III, work on a further two independent studies simultaneously reported association between *DISC1* and neurocognitive function in schizophrenia [169] and normal cognitive aging[168].

Furthermore, by controlling for *DISC1* in a genome-wide scan (IV) evidence was observed that other genes known to interact with *DISC1* may also play a role in the aetiology of schizophrenia. This study highlighted a locus that had previously been linked to bipolar affective disorder and contained the *DISC1* interacting protein *NDE1*. *NDE1*, in turn, was observed to be associated with schizophrenia, being over-transmitted to affected females. Upon closer look at the genomic localisation of other *DISC1* interacting genes it is seen that there is existing genetic evidence for many of them to have an involvement in the aetiology of schizophrenia. This has come through traditional linkage analysis[117], cytogenetic analysis (Millar et al, personal communication) and linkage analysis using endophenotypes of schizophrenia[10], as well as the linkage analysis performed in stage IV. Additionally, a further *DISC1* interacting gene, *FEZ1*, has already been shown to be associated with schizophrenia[189]. At the functional level other *DISC1* binding proteins have been identified as involved in learning and memory functions through animal models, *LIS1* in *Mus musculus* (mouse)[190], and *PDE4B* and *D* in *Drosophila melanogaster* (fruit fly)[191]. It has recently been observed that *DISC1* dynamically interacts with *PDE4B* and *D* in a cAMP dependent manner, providing a plausible mechanistic link to learning and memory (Millar et al. personal communication). Therefore, all the current evidence is highly indicative of a "*DISC1* pathway" in conferring susceptibility to schizophrenia potentially through a role in learning and memory functions. Furthermore, it is interesting to note that the complex in which *LIS1* and the *NDE1* homolog *NDEL1* are known to function, and in which *NDE1* and *DISC1* are now hypothesised to function, in a pathway regulated through signalling of another schizophrenia candidate gene, *RELN*. Such indications of functional pathways in the aetiology of schizophrenia allow for more to be concluded and hypothesised for future analysis on how the normal workings of the human brain are disrupted in creating such a devastating disorder.

6 Concluding remarks and future prospects

Schizophrenia is a major debilitating mental disorder that affects approximately 1% of the worldwide population, and impacts upon the affected individual's family members, communities, and the health care and welfare systems. Genetic analysis of schizophrenia has continued for over twenty years, yet little progress has been made when compared to what was expected of the field. With many researchers into psychiatric genetics coming to be more pragmatic as it has become clearer that major genes probably do not exist for mental illness, and that these disorders are genetically complex and heterogeneous. Yet recently a number of strong candidate genes have been identified that are once again providing hope to the field.

Since the start of this study in September 2000 the 1q42 locus has gone from being just another region linked to schizophrenia, to a locus containing one of the hottest candidate genes for a role in the aetiology of this complex disorder. Primarily this transition has been due to the genetic findings presented here. However, the independent replication of linkage and association to this region, and the dissection of the protein function of *DISC1* have contributed immensely to making this transition all the more credible.

Yet it is Dysbindin (*DTNBP1*), and Neuregulin (*NRG1*) that can currently be considered to be the strongest candidate genes, with many independent replications after their first identification in 2002. Yet despite the convincing genetic evidence for all these genes it is still not possible to say conclusively if they function in causing susceptibility to schizophrenia. Efforts have now generally turned towards the identification of functional risk variations within these genes, when what is probably needed is a more comprehensive approach to the analysis. This would use numerous approaches to identify the role of the genes in the disorder, and would incorporate many fields of expertise, including animal studies, cellular expression analysis, as well as further genetic analysis. Once putative genetic risk variants are identified it will be possible to start to model how that genetic variation affects on schizophrenia when combined with environmental risks, putatively identifying risk factors that must either both occur or interact in order to create a risk to the development of schizophrenia.

In the study presented here stage IV represents the further genetic analysis and presents potentially the most interesting piece of evidence for *DISC1* to be involved in the aetiology of schizophrenia. In this stage, it was found that by controlling for a risk haplotype of *DISC1* a region on chromosome 16 provided evidence of linkage that then led to the identification of association between an allelic haplotype of the *NDE1* gene with schizophrenia. The proteins of *NDE1*

and *DISC1* are known to interact, potentially in the cellular processes governing neuronal migration. Further *in silico* analysis found that a number of other *DISC1* binding proteins are genetically located in regions previously identified as susceptibility regions for schizophrenia, with a number of these genes having animal models that would implicate them in cognitive functions relevant to those identified as associating with *DISC1*. This directly implicates a *DISC1* "pathway" in the aetiology of schizophrenia, which when analysed together, and taking into consideration their biological interactions, could provide a more comprehensive indication of how all these genes create to induce susceptibility to schizophrenia. The role of many of these proteins in neuronal migration would also mean that they should be studied in combination with environmental risk factors affecting during the time that such migration occurs, principally prenatal exposure to infections.

It is this that psychiatric genetic analysis should look to next, not the identification of presumed unrelated genes that individually associate with schizophrenia, but to the identification of pathways that have a determinable role in the normal functioning of the human brain, through which it is possible to conclude upon how the disrupted pathway leads to susceptibility to schizophrenia. Such identified pathways will lead to a more comprehensive understanding of the biological aetiology that underlies the disorder. This has the potential not just to aide understanding of schizophrenia but also many other mental illnesses that potentially have overlapping or similar aetiologies. This would pave the way for new medications and treatments for these debilitating mental disorders, by being able to target specific treatments to specific individuals having specific deficits rather than relying on the same drugs that have wide-ranging non-specific affects, used for all individuals.

7 Acknowledgements

This study was carried out at the Department of Molecular Medicine of the National Public Health Institute, Helsinki. The former head of the institute, Jussi Huttunen, and the new director, Pekka Puska, are acknowledged for providing excellent research facilities. This study was financially supported by Millennium Pharmaceuticals Inc, Wyeth Pharmaceuticals Inc, The Finnish Cultural Foundation's Paavo Koskinen, and Aili and Paul Pennanen grants, and Helsinki Biomedical Graduate School.

This work has been carried out under the supervision of Leena Peltonen-Palotie, who early on was skilfully able to realize that I would respond best to encouragement, and managed to get me to produce my best work. Tiina Paunio is thanked for being a co-supervisor, and in providing me with the motivation and the skills needed to complete this work. I would also like to exceptionally thank Jesper Ekelund for his many invaluable contributions to this project and to Teppo Varilo who has always gone out of his way in helping whenever he could.

I wish to express my appreciation to Hasse Karlsson and Maija Wessman for reviewing this thesis. Their constructive comments and criticism led to some important changes that improved this work. Anna-Elina Lehesjoki, Hasse Karlsson and Kimmo Kontula are thanked for participating in my graduate school thesis committee.

I would like to offer some belated thanks to a number of people in my working life, without whom I never would have got this far. Frank Butler first introduced me to genetics at Strode's College and his enthusiasm and encouragement, upon realizing that I was good, guided me to change my career choice from general biology to genetics. Barbara Daniels was also a tireless supply of encouragement while I studied at King's College in London, if it wasn't for her I might have taken my official supervisors advice and quit science for something less demanding. For my first taste of real science David Collier gets great thanks. In practice the work was a lesson in how to work in a laboratory, in reality it was the start to my work on the genetics of schizophrenia. Last but by no means least, Linzy Hill has my undying gratitude for reaffirming my faith in science and scientists, and for giving me the platform from which I was able to make the leap to Finland and to working for a PhD.

I would like to acknowledge the secretaries, Sari Kivikko, Tuija Svahnbäck, Sari Mustala, Helena Knuuttila, Sami Mustala, and Mika Kivimäki who have always been of assistance when I most needed. Teemu Perheentupa, and Jari Raikko, are thanked for their kind assistance with numerous computer difficulties.

Major thanks go to everybody in the “schizo/psycho” group for a great working atmosphere, primarily, Jesper Ekelund, Joni Turunen and Eva-Maria Pylkkö (Emma P.). Jesper, by diligently explaining difficult concepts and by being so involved in this and the other schizophrenia projects, has acted as a good role model in my professional life. Joni started at around the same time as me on the road to getting a PhD in this group, and over the years we have helped each other immensely both with work and the hard times that come from doing such research at this level. Emma P. as one of a number of newer students in this group, has willingly accepted the information and help I have been able to pass on, and I look forward to seeing the discoveries that she will eventually make. The rest of the group are warmly thanked for sharing an office, and for the discussions we have shared over the years, these are Anu-Maria, Emma N., Heidi, Helena, Johanna, Jenny, Juho, Karola, Katri, Liisa A, Liisa T, Marika, Marjo, Minttu M., Olli, Perttu, Pia, Satu, Sisko, and Tero Y. I would also like to thank those individuals who work on the clinical analysis of schizophrenia, for their great collaborations and patient explanations, especially Jouko Lönnqvist, Timo Partonen, and Annamari Tuulio-Henriksson, but also Jari, Marjut, Ritva, and Tuula. Those that do not work on schizophrenia but have contributed their knowledge in helping me approach this problem are much appreciated, Jaana, Kaisa, Markus, Riika, and Tero H.

Carina von Schantz (Tintti) is exceptionally thanked for taking me under her wing as a friend from my early days at KTL, leading me into the group of people I would come to call my friends. That group evolved into, Tintti, Annina, Emma P. Jesper, Joni, and Pekka, their friendship outside of work has been an immense help in my acclimatisation to life in Finland. Additionally, many others have participated to the great general atmosphere within the KTL Molecular Medicine Department, Denis, Elina P, Elina S., Greg, Johannes, Jussi, Kati, Kirsi, Mervi, Mikko, Minna, Minttu J., Mira, Nabil, Niklas, Päivi, Roxy, Sarah, Siv, Tony, and Ville.

Unending gratitude goes to my parents Kay and John, for their boundless love and support, despite the increased physical distance. It was this support which filled my bedroom with encyclopaedias and wildlife books, and encouraged my already sky high enthusiasm for the natural world. Thanks too goes to my brothers Richard and Kenneth, and their wives Hilary and Naomi, again for their support in my moving countries, but also for not letting me get a word in during our conversations and thereby providing me with the time to think more. Big thanks go to my nephews Michael and Joseph, who through videos and visits have provided great light relief. I am also totally indebted in thanks to Henna’s parents Martti and Anna-Liisa, and her sister Mirkka, for being more than ready to take me in and treat me and as member of the family from my very first steps

from the airport. Life would have been so much harder without a loving family network close at hand.

The most thanks are due to Henna, you were my reason for coming to Helsinki, and you have given me the strength and will that sees me through.

William Hennah
Helsinki 2005

8 References

1. Harrison, P. J. & Weinberger, D. R. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol Psychiatry* **10**, 40-68 (2005).
2. Lewis, C. M. et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* **73**, 34-48 (2003).
3. St Clair, D. et al. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet* **336**, 13-16 (1990).
4. Sachs, N. A. et al. A frameshift mutation in Disrupted in Schizophrenia 1 in an American family with schizophrenia and schizoaffective disorder. *Mol Psychiatry* (2005).
5. Hwu, H. G., Liu, C. M., Fann, C. S., Ou-Yang, W. C. & Lee, S. F. Linkage of schizophrenia with chromosome 1q loci in Taiwanese families. *Mol Psychiatry* **8**, 445-452 (2003).
6. Hodgkinson, C. A. et al. Disrupted in Schizophrenia 1 (DISC1): Association with Schizophrenia, Schizoaffective Disorder, and Bipolar Disorder. *Am J Hum Genet* **75**, 862-872 (2004).
7. Ekelund, J. et al. Chromosome 1 Loci in Finnish Schizophrenia Families. *Human Molecular Genetics* **10**, 1611-1617 (2001).
8. Curtis, D. et al. Genome scan of pedigrees multiply affected with bipolar disorder provides further support for the presence of a susceptibility locus on chromosome 12q23-q24, and suggests the presence of additional loci on 1p and 1q. *Psychiatr Genet* **13**, 77-84 (2003).
9. Thomson, P. A. et al. Association between the TRAX/DISC locus and both bipolar disorder and schizophrenia in the Scottish population. *Mol Psychiatry* (2005).
10. Paunio, T. et al. Search for cognitive trait components of schizophrenia reveals a locus for verbal learning and memory on 4q and for visual working memory on 2q. *Hum Mol Genet* **13**, 1693-1702 (2004).
11. Kendler, K. S. et al. The Roscommon Family Study. I. Methods, diagnosis of probands, and risk of schizophrenia in relatives. *Arch Gen Psychiatry* **50**, 527-540 (1993).
12. Cannon, T. D., Kaprio, J., Lönnqvist, J., Huttunen, M. & Koskenvuo, M. The Genetic Epidemiology of Schizophrenia in a Finnish Twin Cohort. *Arch Gen Psychiatry* **55**, 67-74 (1998).
13. Bassett, A. S., McGillivray, B. C., Jones, B. D. & Pantzar, J. T. Partial trisomy chromosome 5 cosegregating with schizophrenia. *Lancet* **1**, 799-801 (1988).
14. Sherrington, R. et al. Localization of a susceptibility locus for schizophrenia on chromosome 5. *Nature* **336**, 164-167 (1988).
15. Tienari, P. et al. The Finnish adoptive family study of schizophrenia. Implications for family research. *Br J Psychiatry Suppl* **23**, 20-26 (1994).

16. Pastinen, T. et al. Dissecting a population genome for targeted screening of disease mutations. *Hum Mol Genet* **10**, 2961-72 (2001).
17. Norio, R., Nevanlinna, H. R. & Perheentupa, J. Hereditary diseases in Finland; rare flora in rare soul. *Annals of Clinical Research* **5**, 109-141 (1973).
18. Peltonen, L., Jalanko, A. & Varilo, T. Molecular Genetics of the Finnish Disease Heritage. *Human Molecular Genetics* **8**, 1913-1923 (1999).
19. Peltonen, L., Palotie, A. & Lange, K. Use of population isolates for mapping complex traits. *Nat Rev Genet* **1**, 182-90. (2000).
20. Boehnke, M. A look at linkage disequilibrium. *Nat Genet* **25**, 246-7 (2000).
21. Gottesman, I. I. *Schizophrenia genesis* (eds. Atkinson, R., Lindzey, G. & Thompson, R.) (W. H. Freeman and Company, New York, 1991).
22. Schwab, S. G. et al. Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nat Genet* **11**, 325-327 (1995).
23. Schwab, S. G. et al. Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. *Am J Hum Genet* **72**, 185-90. (2003).
24. Straub, R. E. et al. Genetic Variation in the 6p22.3 Gene DTNBP1, the Human Ortholog of the Mouse Dysbindin Gene, is Associated with Schizophrenia. *Am. J. Hum. Genet.* **71**, 337-348 (2002).
25. Straub, R. E. et al. A potential vulnerability locus for schizophrenia on chromosome 6p24- 22: evidence for genetic heterogeneity. *Nat Genet* **11**, 287-293 (1995).
26. Wang, S. et al. Evidence for a susceptibility locus for schizophrenia on chromosome 6pter-p22. *Nat Genet* **10**, 41-6 (1995).
27. Williams, N. M. et al. Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1). *Arch Gen Psychiatry* **61**, 336-44 (2004).
28. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, fourth edition (DSM-IV).* (American Psychiatric Press, Inc., Washington DC, 1994).
29. World Health Organization. *International classification of diseases, 10th ed.* (World Health Organization, Geneva, 1994).
30. Berrettini, W. H. Susceptibility loci for bipolar disorder: overlap with inherited vulnerability to schizophrenia. *Biol Psychiatry* **47**, 245-51 (2000).
31. Impagnatiello, F. et al. A decrease of reelin expression as a putative vulnerability factor in schizophrenia. *Proc Natl Acad Sci U S A* **95**, 15718-23 (1998).

32. Ekelund, J. et al. Genome-Wide Scan for Schizophrenia in the Finnish Population: Evidence for a Locus on Chromosome 7q22. *Human Molecular Genetics* **9**, 1049-1057 (2000).
33. Egan, M. F. et al. Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. *Proc Natl Acad Sci U S A* **101**, 12604-9 (2004).
34. Fujii, Y. et al. Positive associations of polymorphisms in the metabotropic glutamate receptor type 3 gene (GRM3) with schizophrenia. *Psychiatr Genet* **13**, 71-6 (2003).
35. Torrey, E. F. Prevalence studies in schizophrenia. *Br J Psychiatry* **150**, 598-608 (1987).
36. Hovatta, I. et al. Schizophrenia in the Genetic Isolate of Finland. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **74**, 353-360 (1997).
37. Haukka, J., Suvisaari, J., Varilo, T. & Lonnqvist, J. Regional variation in the incidence of schizophrenia in Finland: a study of birth cohorts born from 1950 to 1969. *Psychol Med* **31**, 1045-53 (2001).
38. McGrath, J. J. et al. The fertility and fecundity of patients with psychoses. *Acta Psychiatr Scand* **99**, 441-446 (1999).
39. Brune, M. Schizophrenia-an evolutionary enigma? *Neurosci Biobehav Rev* **28**, 41-53 (2004).
40. Blouin, J. L. et al. Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* **20**, 70-73 (1998).
41. Pulver, A. E. et al. The Johns Hopkins University Collaborative Schizophrenia Study: an epidemiologic-genetic approach to test the heterogeneity hypothesis and identify schizophrenia susceptibility genes. *Cold Spring Harb Symp Quant Biol* **61**, 797-814 (1996).
42. Stefansson, H. et al. Neuregulin 1 and Susceptibility to Schizophrenia. *Am. J. Hum. Genet.* **71**, 877-892 (2002).
43. Stefansson, H. et al. Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. *Am J Hum Genet* **72**, 83-7. (2003).
44. Tang, J. X. et al. Polymorphisms within 5' end of the Neuregulin 1 gene are genetically associated with schizophrenia in the Chinese population. *Mol Psychiatry* **9**, 11-2 (2004).
45. Hafner, H. Gender differences in schizophrenia. *Psychoneuroendocrinology* **28 Suppl 2**, 17-54 (2003).
46. Kendler, K. S. & Walsh, D. Gender and schizophrenia. Results of an epidemiologically-based family study. *Br J Psychiatry* **167**, 184-192 (1995).
47. Castle, D. J., Abel, K., Takei, N. & Murray, R. M. Gender differences in schizophrenia: hormonal effect or subtypes? *Schizophr Bull* **21**, 1-12 (1995).

48. Mowry, B. J. et al. Second stage of a genome scan of schizophrenia: study of five positive regions in an expanded sample. *Am J Med Genet* **96**, 864-9 (2000).
49. Levinson, D. F. et al. Genome scan of schizophrenia. *Am J Psychiatry* **155**, 741-50 (1998).
50. Fallin, M. D. et al. Genomewide Linkage Scan for Schizophrenia Susceptibility Loci among Ashkenazi Jewish Families Shows Evidence of Linkage on Chromosome 10q22. *Am J Hum Genet* **73**, 601-11 (2003).
51. Ewald, H. et al. Search for a shared segment on chromosome 10q26 in patients with bipolar affective disorder or schizophrenia from the Faroe Islands. *Am J Med Genet* **114**, 196-204 (2002).
52. Palmer, C. G. et al. RHD maternal-fetal genotype incompatibility increases schizophrenia susceptibility. *Am J Hum Genet* **71**, 1312-9 (2002).
53. Cannon, M., Jones, P. B. & Murray, R. M. Obstetric complications and schizophrenia: historical and meta-analytic review. *Am J Psychiatry* **159**, 1080-92 (2002).
54. Chumakov, I. et al. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc Natl Acad Sci U S A* **99**, 13675-80. (2002).
55. Brown, A. S. et al. Serologic evidence of prenatal influenza in the etiology of schizophrenia. *Arch Gen Psychiatry* **61**, 774-80 (2004).
56. Brown, A. S., Cohen, P., Greenwald, S. & Susser, E. Nonaffective psychosis after prenatal exposure to rubella. *Am J Psychiatry* **157**, 438-43 (2000).
57. Barr, C. E., Mednick, S. A. & Munk-Jorgensen, P. Exposure to influenza epidemics during gestation and adult schizophrenia. A 40-year study. *Arch Gen Psychiatry* **47**, 869-874 (1990).
58. Mortensen, P. B. et al. Effects of family history and place and season of birth on the risk of schizophrenia. *N Engl J Med* **340**, 603-8 (1999).
59. Arseneault, L. et al. Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study. *BMJ* **325**, 1212-3 (2002).
60. Kirov, G. et al. Strong evidence for association between the dystrobrevin binding protein 1 gene (DTNBP1) and schizophrenia in 488 parent-offspring trios from Bulgaria. *Biol Psychiatry* **55**, 971-5 (2004).
61. Morris, D. W. et al. No evidence for association of the dysbindin gene [DTNBP1] with schizophrenia in an Irish population-based study. *Schizophr Res* **60**, 167-72 (2003).
62. Tang, J. X. et al. Family-based association study of DTNBP1 in 6p22.3 and schizophrenia. *Mol Psychiatry* **8**, 1008 (2003).
63. Van Den Bogaert, A. et al. The DTNBP1 (Dysbindin) Gene Contributes to Schizophrenia, Depending on Family History of the Disease. *Am J Hum Genet* **73**, 1438-1443 (2003).

64. van den Oord, E. J. et al. Identification of a high-risk haplotype for the dystrobrevin binding protein 1 (DTNBP1) gene in the Irish study of high-density schizophrenia families. *Mol Psychiatry* **8**, 499-510 (2003).
65. Gottesman, I. I. & Gould, T. D. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* **160**, 636-645 (2003).
66. Tuulio-Henriksson, A. et al. Heritability and number of quantitative trait loci of neurocognitive functions in families with schizophrenia. *Am J Med Genet* **114**, 483-490 (2002).
67. Williams, N. M. et al. Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia. *Mol Psychiatry* **8**, 485-7 (2003).
68. Yang, J. Z. et al. Association study of neuregulin 1 gene with schizophrenia. *Mol Psychiatry* **8**, 706-9 (2003).
69. McGuffin, P., Farmer, A. E. & Harvey, I. A Polydiagnostic Application of Operational Criteria in Studies of Psychotic Illness. Development and Reliability of the OPCRIT System. *Arch Gen Psychiatry* **48**, 764-770 (1991).
70. Williams, J., Farmer, A. E., Ackenheil, M., Kaufmann, C. A. & McGuffin, P. A Multicentre Inter-Rater Reliability Study Using OPCRIT Computerized Diagnostic System. *Psychological Medicine* **26**, 775-783 (1996).
71. Freedman, R. et al. Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. *Proc Natl Acad Sci U S A* **94**, 587-592 (1997).
72. Freedman, R. et al. Inhibitory neurophysiological deficit as a phenotype for genetic investigation of schizophrenia. *Am J Med Genet* **97**, 58-64. (2000).
73. Tsuang, M. Schizophrenia: genes and environment. *Biol Psychiatry* **47**, 210-20. (2000).
74. Cardno, A. G. et al. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry* **56**, 162-168 (1999).
75. Kendler, K. S., Gruenberg, A. M. & Kinney, D. K. Independent diagnoses of adoptees and relatives as defined by DSM-III in the provincial and national samples of the Danish Adoption Study of Schizophrenia. *Arch Gen Psychiatry* **51**, 456-468 (1994).
76. Kety, S. S. et al. Mental illness in the biological and adoptive relatives of schizophrenic adoptees. Replication of the Copenhagen Study in the rest of Denmark. *Arch Gen Psychiatry* **51**, 442-455 (1994).
77. Brzustowicz, L. M., Hodgkinson, K. A., Chow, E. W. C., Honer, W. G. & Bassett, A. S. Location of a Major Susceptibility Locus for Familial Schizophrenia on Chromosome 1q21-q22. *Science* **288**, 678-682 (2000).

78. Brzustowicz, L. M., Hayter, J. E., Hodgkinson, K. A., Chow, E. W. & Bassett, A. S. Fine mapping of the schizophrenia susceptibility locus on chromosome 1q22. *Hum Hered* **54**, 199-209 (2002).
79. Chowdari, K. V. et al. Association and linkage analyses of RGS4 polymorphisms in schizophrenia. *Human Molecular Genetics* **11**, 1373-1380 (2002).
80. Gurling, H. M. et al. Genomewide Genetic Linkage Analysis Confirms the Presence of Susceptibility Loci for Schizophrenia, on Chromosomes 1q32.2, 5q33.2, and 8p21-22 and Provides Support for Linkage to Schizophrenia, on Chromosomes 11q23.3-24 and 20q12.1-11.23. *American Journal of Human Genetics* **68**, 661-673. (2001).
81. Blackwood, D. H. R. et al. Schizophrenia and Affective Disorders-Cosegregation with a Translocation at Chromosome 1q42 that Directly Disrupts Brain-Expressed Genes: Clinical and P300 Findings in a Family. *American Journal of Human Genetics* **69**, 428-433 (2001).
82. Callicott, J. H. et al. Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proc Natl Acad Sci U S A* **102**, 8627-32 (2005).
83. Schwab, S. G. et al. Evidence suggestive of a locus on chromosome 5q31 contributing to susceptibility for schizophrenia in German and Israeli families by multipoint affected sib-pair linkage analysis. *Mol Psychiatry* **2**, 156-160 (1997).
84. Straub, R. E., MacLean, C. J., O'Neill, F. A., Walsh, D. & Kendler, K. S. Support for a possible schizophrenia vulnerability locus in region 5q22- 31 in Irish families. **2**, 148-155 (1997).
85. Lo, W. S. et al. Association of SNPs and haplotypes in GABAA receptor beta2 gene with schizophrenia. *Mol Psychiatry* **9**, 603-8 (2004).
86. Ambrosio, A. M. et al. A linkage study between the GABAA beta2 and GABAA gamma2 subunit genes and major psychoses. *CNS Spectr* **10**, 57-61 (2005).
87. Byerley, W. et al. Evidence for a 5q31 schizophrenia locus in a large multiplex kindred from Palau, Micronesia. *Molecular Psychiatry* **4**, 4 (1999).
88. Paunio, T. et al. Genome-Wide Scan in a Nationwide Study Sample of Schizophrenia Families in Finland Reveals Susceptibility Loci on Chromosomes 2q and 5q. *Human Molecular Genetics* **10**, 3037-3048 (2001).
89. Moises, H. W. et al. An international two-stage genome-wide search for schizophrenia susceptibility genes. **11**, 321-324 (1995).
90. Maziade, M. et al. 6p24-22 region and major psychoses in the Eastern Quebec population. Le Groupe IREP. *Am J Med Genet* **74**, 311-8 (1997).

91. Kaufmann, C. A. et al. NIMH Genetics Initiative Millenium Schizophrenia Consortium: linkage analysis of African-American pedigrees. *Am J Med Genet* **81**, 282-9 (1998).
92. Petryshen, T. L. et al. Genetic investigation of chromosome 5q GABA(A) receptor subunit genes in schizophrenia. *Mol Psychiatry*, in press (2005).
93. Pimm, J. et al. The Epsin 4 gene on chromosome 5q, which encodes the clathrin-associated protein enthoprotin, is involved in the genetic susceptibility to schizophrenia. *Am J Hum Genet* **76**, 902-7 (2005).
94. Kendler, K. S. et al. Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish Study of High-Density Schizophrenia Families. *Am J Psychiatry* **153**, 1534-1540 (1996).
95. Shaw, S. H. et al. A genome-wide search for schizophrenia susceptibility genes. *Am J Med Genet* **81**, 364-376 (1998).
96. Brzustowicz, L. M. et al. Linkage of familial schizophrenia to chromosome 13q32. *Am J Hum Genet* **65**, 1096-1103 (1999).
97. Gill, M. et al. A combined analysis of D22S278 marker alleles in affected sib-pairs: support for a susceptibility locus for schizophrenia at chromosome 22q12. Schizophrenia Collaborative Linkage Group (Chromosome 22). *Am J Med Genet* **67**, 40-5 (1996).
98. Coon, H. et al. Analysis of chromosome 22 markers in nine schizophrenia pedigrees. *Am J Med Genet* **54**, 72-79 (1994).
99. Karayiorgou, M. et al. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci U S A* **92**, 7612-6 (1995).
100. Pulver, A. E. et al. Follow-up of a report of a potential linkage for schizophrenia on chromosome 22q12-q13.1: Part 2. *Am J Med Genet* **54**, 44-50 (1994).
101. Pulver, A. E. et al. Sequential strategy to identify a susceptibility gene for schizophrenia: report of potential linkage on chromosome 22q12-q13.1: Part 1. *Am J Med Genet* **54**, 36-43 (1994).
102. Li, T. et al. Preferential transmission of the high activity allele of COMT in schizophrenia. *Psychiatr Genet* **6**, 131-3 (1996).
103. Liu, H. et al. Genetic variation in the 22q11 locus and susceptibility to schizophrenia. *Proc Natl Acad Sci U S A* **99**, 16859-64 (2002).
104. Liu, H. et al. Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. *Proc Natl Acad Sci U S A* **99**, 3717-22 (2002).
105. Hovatta, I. et al. A Genomewide Screen for Schizophrenia Genes in an Isolated Finnish Subpopulation, Suggesting Multiple Susceptibility Loci. *American Journal of Human Genetics* **65**, 1114-1124 (1999).

106. Gasperoni, T. L. et al. Genetic linkage and association between chromosome 1q and working memory function in schizophrenia. *Am J Med Genet* **116B**, 8-16 (2003).
107. Millar, J. K. et al. Disruption of Two Novel Genes by a Translocation Co-Segregating with Schizophrenia. *Human Molecular Genetics* **9**, 1415-1423 (2000).
108. Devon, R. S. et al. Identification of Polymorphisms within Disrupted In Schizophrenia 1 and Disrupted In Schizophrenia 2, and an Investigation of their Association with Schizophrenia and Bipolar Affective Disorder. *Psychiatric Genetics* **11**, 71-78 (2001).
109. Levinson, D. F. et al. No major schizophrenia locus detected on chromosome 1q in a large multicenter sample. *Science* **296**, 739-41 (2002).
110. Macgregor, S. et al. Is schizophrenia linked to chromosome 1q? *Science* **298**, 2277 (2002).
111. Bassett, A. S., Chow, E. W., Vieland, V. J. & Brzustowicz, L. Is schizophrenia linked to chromosome 1q? *Science* **298**, 2277 (2002).
112. Weeks, D. E. & Lathrop, G. M. Polygenic disease: methods for mapping complex disease traits. *Trends Genet* **11**, 513-9 (1995).
113. Strachan, T. & Read, A. P. *Human molecular genetics 3* (Garland,, London ; New York, 2004).
114. Ott, J. *Analysis of human genetic linkage* (Johns Hopkins University Press, Baltimore, USA, 1999).
115. Klein, R. J. et al. Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-9 (2005).
116. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**, 241-247 (1995).
117. Sklar, P. Linkage analysis in psychiatric disorders: the emerging picture. *Annu Rev Genomics Hum Genet* **3**, 371-413 (2002).
118. Varilo, T. et al. The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum Mol Genet* **12**, 51-9 (2003).
119. Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. & Lander, E. S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* **58**, 1347-1363 (1996).
120. Sobel, E. & Lange, K. Descent Graphs in Pedigree Analysis: Applications to Haplotyping, Location Scores, and Marker-Sharing Statistics. *American Journal of Human Genetics* **58**, 1323-1337 (1996).
121. Schaid, D. J., McDonnell, S. K., Wang, L., Cunningham, J. M. & Thibodeau, S. N. Caution on pedigree haplotype inference with software that assumes linkage equilibrium. *Am J Hum Genet* **71**, 992-5 (2002).

122. Rohde, K. & Fuerst, R. Haplotyping and estimation of haplotype frequencies for closely linked biallelic multilocus genetic phenotypes including nuclear family information. *Hum Mutat* **17**, 289-95 (2001).
123. O'Connell, J. R. & Weeks, D. E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* **63**, 259-266 (1998).
124. Lange, K., Weeks, D. & Boehnke, M. Programs for Pedigree Analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol* **5**, 471-2 (1988).
125. Clayton, D. A Generalization of the Transmission/Disequilibrium Test for Uncertain-Haplotype Transmission. *American Journal of Human Genetics* **65**, 1170-1177 (1999).
126. Horvath, S. et al. Family Based Tests for Associating Haplotypes with General Phenotype Data: Application to Asthma Genetics. *Genet Epidemiol* **26**, 61-69 (2004).
127. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-89 (2001).
128. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30**, 97-101 (2002).
129. Abecasis, G. R., Cardon, L. R. & Cookson, W. O. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* **66**, 279-292 (2000).
130. Cannon, T. D. et al. Neuropsychological functioning in siblings discordant for schizophrenia and healthy volunteers. *Arch Gen Psychiatry* **51**, 651-61 (1994).
131. Lee, W. C. Testing for candidate gene linkage disequilibrium using a dense array of single nucleotide polymorphisms in case-parents studies. *Epidemiology* **13**, 545-51. (2002).
132. Heller, J. *Catch-22* (Simon and Schuster, New York,, 1961).
133. van den Oord, E. J. & Sullivan, P. F. False discoveries and models for gene discovery. *Trends Genet* **19**, 537-42 (2003).
134. Hovatta, I. et al. Linkage Analysis of Putative Schizophrenia Gene Candidate Regions on Chromosome 3p, 5q, 6p, 8p, 20p, and 22q in a Population-Based Sampled Finnish Family Set. *Molecular Psychiatry* **3**, 452-457 (1998).
135. Lichtermann, D., Hovatta, I., Terwilliger, J. D., Peltonen, L. & Lönnqvist, J. Concordance for Sex and the Pseudoautosomal Gene Hypothesis Revisited: No Evidence of Increased Sex Concordance in a Nationwide Finnish Sample of Siblings with Paternally Derived Schizophrenia. *American Journal of Psychiatry* **155**, 1365-1375 (1998).

136. Kendler, K. S. et al. The Roscommon Family Study. III. Schizophrenia-related personality disorders in relatives. *Arch Gen Psychiatry* **50**, 781-788 (1993).
137. Arajarvi, R. et al. Clinical phenotype of schizophrenia in a Finnish isolate. *Schizophr Res* **67**, 195-205 (2004).
138. Isohanni, M. et al. A comparison of clinical and research DSM-III-R diagnoses of schizophrenia in a Finnish national birth cohort. Clinical and research diagnoses of schizophrenia. *Social Psychiatry & Psychiatric Epidemiology* **32**, 303-308 (1997).
139. Mäkikyrö, T. et al. Accuracy of register-based schizophrenia diagnoses in a genetic study. *European Psychiatry* **13**, 57-62 (1998).
140. Cannon, T. D. et al. The inheritance of neuropsychological dysfunction in twins discordant for schizophrenia. *Am J Hum Genet* **67**, 369-382. (2000).
141. Tuulio-Henriksson, A. et al. Familial loading associates with impairment in visual span among healthy siblings of schizophrenia patients. *Biol Psychiatry* **54**, 623-628 (2003).
142. Wechsler, D. *Wechsler Memory Scale - Revised (WMS-R), manual.* (San Antonio: Psychological Corporation, Harcourt Brace Jovanovich, Inc., 1987).
143. Delis, D. C., Kramer, J. H., Kaplan, E. & Ober, B. A. *California Verbal Learning Test: Manual, Research Edition.* (San Antonio: Psychological Corporation, Harcourt Brace & Company., 1987).
144. Wechsler, D. *Wechsler Adult Intelligence Scale - Revised (WAIS-R), manual.* (Cleveland: Psychological Corporation, Harcourt Brace Jovanovich, Inc., 1981).
145. Blin, N. & Stafford, D. W. A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Research* **3**, 2303-2308 (1976).
146. Syvänen, A. C., Aalto-Setala, K., Harju, L., Kontula, K. & Soderlund, H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**, 684-692 (1990).
147. Pastinen, T. et al. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res* **10**, 1031-1042 (2000).
148. Jurinke, C., van den Boom, D., Cantor, C. R. & Koster, H. Automated genotyping using the DNA MassArray technology. *Methods Mol Biol* **170**, 103-16 (2001).
149. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A* **81**, 3443-6 (1984).
150. Lathrop, G. M. & Lalouel, J. M. Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet* **36**, 460-5 (1984).

151. Lathrop, G. M., Lalouel, J. M. & White, R. L. Construction of human linkage maps: likelihood calculations for multilocus linkage analysis. *Genet Epidemiol* **3**, 39-52 (1986).
152. Schaffer, A. A., Gupta, S. K., Shriram, K. & Cottingham, R. W., Jr. Avoiding recomputation in linkage analysis. *Hum Hered* **44**, 225-37 (1994).
153. Cottingham, R. W., Jr., Idury, R. M. & Schaffer, A. A. Faster sequential genetic linkage computations. *Am J Hum Genet* **53**, 252-63 (1993).
154. Terwilliger, J. D., Speer, M. & Ott, J. Chromosome-based method for rapid computer simulation in human genetic linkage analysis. *Genet Epidemiol* **10**, 217-24 (1993).
155. Terwilliger, J. D. & Ott, J. *Handbook of human genetic linkage* (Johns Hopkins University Press,, Baltimore, 1994).
156. Hiekkalinna, T., Terwilliger, J. D., Sammalisto, S., Peltonen, L. & Perola, M. AUTOGSCAN: powerful tools for automated genome-wide linkage and linkage disequilibrium analysis. *Twin Res Hum Genet* **8**, 16-21 (2005).
157. Göring, H. H. H. & Terwilliger, J. D. Linkage Analysis in the Presence of Errors IV: Joint Pseudomarker Analysis of Linkage and/or Linkage Disequilibrium on a Mixture of Pedigrees and Singletons when the Mode of Inheritance Cannot be Accurately Specified. *American Journal of Human Genetics* **66**, 1310-1327 (2000).
158. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-5 (2005).
159. Roberts, S. B., MacLean, C. J., Neale, M. C., Eaves, L. J. & Kendler, K. S. Replication of linkage studies of complex traits: an examination of variation in location estimates. *American Journal of Human Genetics* **65**, 876-884 (1999).
160. Castle, D. J., Wessely, S. & Murray, R. M. Sex and schizophrenia: effects of diagnostic stringency, and associations with and premorbid variables. *Br J Psychiatry* **162**, 658-664 (1993).
161. Reik, W. & Walter, J. Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote. *Nat Genet* **27**, 255-6 (2001).
162. Ostrer, H. Genome and Hormones: Gender Differences in Physiology. Invited Review: Sex-Based Differences in Gene Expression. **91**, 2384-2388 (2001).
163. Sharma, T. & Antonova, L. Cognitive function in schizophrenia. Deficits, functional consequences, and future treatment. *Psychiatr Clin North Am* **26**, 25-40 (2003).
164. Millar, J. K., Christie, S. & Porteous, D. J. Yeast two-hybrid screens implicate DISC1 in brain development and function. *Biochem Biophys Res Commun* **311**, 1019-1025 (2003).

165. Brandon, N. J. et al. Disrupted in Schizophrenia 1 and Nudel form a neurodevelopmentally regulated protein complex: implications for schizophrenia and other major neurological disorders. *Mol Cell Neurosci* **25**, 42-55 (2004).
166. Millar, J. K., James, R., Brandon, N. J. & Thomson, P. A. DISC1 and DISC2: discovering and dissecting molecular mechanisms underlying psychiatric illness. *Ann Med* **36**, 367-78 (2004).
167. Ekholm, J. M. et al. Evidence of susceptibility loci on 4q32 and 16p12 for bipolar disorder. *Hum Mol Genet* **12**, 1907-15 (2003).
168. Thomson, P. A. et al. Association between genotype at an exonic SNP in DISC1 and normal cognitive aging. *Neurosci Lett*, in press (2005).
169. Burdick, K. E. et al. DISC1 and neurocognitive function in schizophrenia. *Neuroreport* **16**, 1399-402 (2005).
170. Wynshaw-Boris, A. & Gambello, M. J. LIS1 and dynein motor function in neuronal migration and development. *Genes Dev* **15**, 639-51 (2001).
171. Cahana, A. et al. Targeted mutagenesis of Lis1 disrupts cortical development and LIS1 homodimerization. *Proc Natl Acad Sci U S A* **98**, 6429-34 (2001).
172. Hirotsune, S. et al. Graded reduction of Pafah1b1 (Lis1) activity results in neuronal migration defects and early embryonic lethality. *Nat Genet* **19**, 333-9 (1998).
173. Feng, Y. & Walsh, C. A. Mitotic spindle regulation by Nde1 controls cerebral cortical size. *Neuron* **44**, 279-93 (2004).
174. Seidman, L. J. et al. Left hippocampal volume as a vulnerability indicator for schizophrenia: a magnetic resonance imaging morphometric study of nonpsychotic first-degree relatives. *Arch Gen Psychiatry* **59**, 839-49 (2002).
175. Cannon, T. D. et al. Cortex mapping reveals regionally specific patterns of genetic and disease-specific gray-matter deficits in twins discordant for schizophrenia. *Proc Natl Acad Sci U S A* **99**, 3228-33 (2002).
176. Cannon, T. D. et al. DISC1/TRAX Haplotypes Associate with Schizophrenia, Reduced Prefrontal Gray Matter, and Impaired Short- and Long-Term Memory. *Arch Gen Psychiatry*, in press (2005).
177. Millar, J. K., Christie, S., Semple, C. A. M. & Porteous, D. J. Chromosomal Location and Genomic Structure of the Human Translin-Associated Factor X Gene (TRAX; TSNAX) Revealed by Intergenic Splicing to DISC1, a Gene Disrupted by a Translocation Segregating with Schizophrenia. *Genomics* **67**, 69-77 (2000).
178. Ma, L. et al. Cloning and Characterization of Disc1, the Mouse Ortholog of DISC1 (Disrupted-in-Schizophrenia 1). *Genomics* **80**, 662-72. (2002).

179. Taylor, M. S., Devon, R. S., Millar, J. K. & Porteous, D. J. Evolutionary constraints on the Disrupted in Schizophrenia locus. *Genomics* **81**, 67-77. (2003).
180. James, R. et al. Disrupted in Schizophrenia 1 (DISC1) is a multicompartmentalized protein that predominantly localizes to mitochondria. *Mol Cell Neurosci* **26**, 112-22 (2004).
181. Miyoshi, K. et al. Disrupted-In-Schizophrenia 1, a candidate gene for schizophrenia, participates in neurite outgrowth. *Mol Psychiatry* **8**, 685-694 (2003).
182. Morris, J. A., Kandpal, G., Ma, L. & Austin, C. P. DISC1 (Disrupted-In-Schizophrenia 1) is a centrosome-associated protein that interacts with MAP1A, MIPT3, ATF4/5 and NUDEL: regulation and loss of interaction with mutation. *Hum Mol Genet* **12**, 1591-1608 (2003).
183. Ozeki, Y. et al. Disrupted-in-Schizophrenia-1 (DISC-1): mutant truncation prevents binding to NudE-like (NUDEL) and inhibits neurite outgrowth. *Proc Natl Acad Sci U S A* **100**, 289-294 (2003).
184. Finkenstadt, P. M. et al. Somatodendritic Localization of Translin, a Component of the Translin/TRAX RNA Binding Complex. *Journal of Neurochemistry* **75**, 1754-1762 (2000).
185. Zhang, X. et al. Association study of the DISC1/TRAX locus with schizophrenia in a Japanese population. *Schizophr Res* (2005).
186. Kockelkorn, T. T. et al. Association study of polymorphisms in the 5' upstream region of human DISC1 gene with schizophrenia. *Neurosci Lett* **368**, 41-5 (2004).
187. van Erp, T. G. et al. Hippocampal volumes in schizophrenic twins. *Arch Gen Psychiatry* **61**, 346-53 (2004).
188. Cannon, T. D. et al. Regional gray matter, white matter, and cerebrospinal fluid distributions in schizophrenic patients, their siblings, and controls. *Arch Gen Psychiatry* **55**, 1084-91 (1998).
189. Yamada, K. et al. Association analysis of FEZ1 variants with schizophrenia in Japanese cohorts. *Biol Psychiatry* **56**, 683-90 (2004).
190. Paylor, R. et al. Impaired learning and motor behavior in heterozygous Pafah1b1 (Lis1) mutant mice. *Learn Mem* **6**, 521-37 (1999).
191. Cherry, J. A. & Davis, R. L. A mouse homolog of dunce, a gene important for learning and memory in Drosophila, is preferentially expressed in olfactory receptor neurons. *J Neurobiol* **28**, 102-13 (1995).